

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number
WO 02/099130 A2(51) International Patent Classification⁷: C12Q 1/68

(74) Agent: CAMPBELL, Patrick, John, Henry; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5JJ (GB).

(21) International Application Number: PCT/GB02/02642

(22) International Filing Date: 7 June 2002 (07.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0113907.0 7 June 2001 (07.06.2001) GB

(71) Applicant (for all designated States except US): UNIVERSITY COLLEGE LONDON [GB/GB]; Gower Street, London WC1E 9BT (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): GRIFFITHS, David, John [GB/GB]; Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF (GB). KELLAM, Paul [GB/GB]; Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF (GB). WEISS, Robert, Anthony [GB/GB]; Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette

(54) Title: VIRUS DETECTION USING DEGENERATE PCR PRIMERS

(57) Abstract: A high throughput method for screening a biological sample for unknown viruses, which method comprises: (a) subjecting DNA from the sample to PCR amplification conditions using simultaneously multiple pairs of degenerate primers, wherein each primer binds a sequence that is conserved across members of a family of viruses and each pair of primers selectively directs amplification of sequence of said family; (b) sequencing PCR product obtained in step (a); and (c) comparing the sequence of the PCR product with the sequences in at least one database comprising viral sequences to determine whether the sequence is present in, or absent from, the database, wherein absence of the sequence from the database suggests that the sequence may be from an unknown virus.

VIRUS DETECTION USING DEGENERATE PCR PRIMERS

Field of the invention

The invention relates to a method of detecting new viruses using a high throughput polymerase chain reaction (PCR) assay.

Background of the invention

5 Biological materials can often become contaminated or infected with unidentified organisms. For example, cells grown in tissue culture often exhibit signs of a cytopathic effect consistent with a virus infection but the identity of the virus may not be apparent. Human blood products, such as factor VIII for the treatment of haemophiliacs, can be contaminated with unidentified viruses, as was
10 demonstrated by infection of many haemophiliacs with human immunodeficiency virus in the early 1980s. Similarly, two decades ago 20% of individuals who received transfused blood contracted hepatitis C (Randall, 2001, J Pediatr. Oncol. Nurs. 18(1), 4-15).

Summary of the invention

PCR allows amplification of a specific region of a polynucleotide. The specificity of the reaction is due to the primers which, during the course of PCR, bind to the region to be amplified in a sequence specific manner. Degenerate primers can be designed which amplify sequence from substantially all members of a virus
20 family. Such primers typically bind to nucleotide sequence which is conserved across the virus family. The invention provides a PCR based high throughput screen that uses such degenerate primers for detecting unknown viruses.

In particular, the invention provides a high throughput method for screening a biological sample for unknown viruses, which method comprises

25 (a) subjecting DNA from the sample to PCR amplification conditions using simultaneously multiple pairs of degenerate primers, wherein each primer binds a sequence that is conserved across members of a family of viruses and each pair of primers selectively directs amplification of sequence of said family;

-2-

(b) sequencing PCR product obtained in step (a); and

- (c) comparing the sequence of the PCR product with the sequences in at least one database comprising viral sequences to determine whether the sequence is present in, or absent from, the database, wherein absence of the sequence from the database suggests that the sequence may be from an unknown virus.

Detailed description of the invention

General description

- There are a number of human diseases in which unidentified viruses are thought to play a causative role. For example, unidentified viruses are believed to play a role in cancers such as leukaemia, autoimmune diseases such as rheumatic disease, cardiovascular diseases such as dilated cardiomyopathy and Kawasaki disease, and prostatitis (zur Hausen 2001 The Lancet 357, 381-384; Greaves 1997 The Lancet 349, 344-349; Rowley and Shulman 1998 Clinical Microbiology Reviews 11(3), 405-414; Kawai 1999 Circulation 99, 1091-1100; and Dominigue and Hellstrom 1998 Clinical Microbiology Reviews 11(4), 604-613). Particles resembling retroviruses have been reported in affected tissue from patients with psoriasis, Sjögren's syndrome and rheumatoid arthritis (Iversen 1990 J. Invest. Dermatol. 90, 41S-3S; Garry et al 1990 Science 250, 1127-9; Yamano et al 1997 J. Clin. Pathol. 50, 223-30; and Stransky et al 1993 Br. J. Rheumatol. 32, 1044-8). The invention provides a way of screening for the viruses which may cause or contribute to such diseases. Once identified, the viruses may be used as a target for developing diagnostic tests for, or therapies against, the diseases.

- The method of the invention is based on obtaining sequences from viruses so that they can be compared with known viral sequences to determine whether they are from novel viruses. The sequences of the novel viruses are amplified using PCR primers which recognise sequences which are conserved (similar/homologous) in known members of virus families. The primers direct amplification of sequence between the conserved regions to give a PCR product whose sequence can be compared with that of known viruses.

The biological sample which is screened may be any sample susceptible to

-3-

infection by a virus. It may, for example, be a tissue culture sample (e.g. tissue culture supernatant), or a sample of animal (including human) or plant material. In a particularly preferred embodiment the invention is directed to the identification of unknown human viruses, and in this case the sample will generally be derived from one or more humans. A sample derived from a human or animal may be from a range of tissue and fluid types, for example blood serum, seminal fluid, breast milk, saliva, cerebrospinal fluid, urine, bile, bronchial lavage fluid, nasal secretion, eye secretion or vaginal wash.

Before the sample is subject to PCR it may be processed. In one embodiment the virus material in the sample is concentrated, for example by ultracentrifugation. The virus material may also be purified in a manner which increases the content of viral nucleic acid relative to non-viral nucleic acid. For example the viral nucleic acid may be concentrated by

centrifuging the biological sample under conditions such that cell debris is pelleted and virus particles remain in the supernatant; collecting the supernatant; and centrifuging the supernatant under conditions such that virus particles are pelleted.

The initial centrifugation to pellet the cell debris may, for example, be carried out at 100 to 10,000 g, preferably from 1000 to 10,000 g. The subsequent centrifugation to pellet the virus particles is carried out at a higher g force, for example 50,000 to 500,000 g, preferably about 100,000 g.

The purification of viral nucleic acid may include a step of treating a suspension comprising the virus with a nuclease so as to digest extraneous nucleic acid, wherein the viral nucleic acid is protected from digestion by viral coat or core protein. The nuclease is preferably a non sequence-specific nuclease which digests DNA and/or RNA, for example micrococcal nuclease S7 (Roche Molecular Biochemicals, Catalogue 107 921).

The processing may also comprise a nucleic acid purification, such as phenol/chloroform nucleic acid purification or the use of a column which selectively binds nucleic acid. In one embodiment purification is carried out using a Qiagen™

column.

Typically processing of the sample increases the purity of the virus nucleic acid present in the sample (for example leading to an increase in concentration of 2-fold to 1000-fold of viral nucleic acid).

- 5 The processing of the sample may comprise the reverse transcription of viral RNA in the sample to DNA, i.e. RNA from the unknown virus is processed to produce the equivalent (such as the same or a complementary) DNA sequence. Thus the DNA which is subject to PCR conditions may be cDNA. This is required when the unknown virus has an RNA genome. Thus the processing may comprise reverse
- 10 transcription of the RNA to produce a complementary DNA strand and then optionally synthesising a second DNA strand before carrying out PCR. This can be achieved by using a primer which directs initiation at random sequences in a reverse transcription reaction and then in a second strand synthesis reaction.

- Random reverse transcription may be directed using a primer which directs
- 15 initiation of DNA synthesis at random sequences. Such a primer may be made by synthesising it so that it contains a random sequence, for example a sequence of at least 6 consecutive nucleotides (e.g. from 6 to 20 nucleotides) wherein each nucleotide may be any of the four possible natural nucleotides, i.e. A, T, C or G. In other words, such a primer contains a sequence NNNNNN wherein each N is A, T, C
- 20 or G.

- In one embodiment a "single tube system" is used for the reverse transcription and then PCR with the multiple pairs of degenerate primers. In such a system the sample (typically after being processed) is added to a mixture of reagents which allow both reverse transcription and PCR to occur. Thus typically the mixture
- 25 will comprise both a reverse transcriptase and a thermostable DNA polymerase. The mixture may comprise the Titan™ reagents from Roche Molecular Biochemicals™ (cat no. 1855476) which uses the avian myeloblastosis virus reverse transcriptase and a Pwo (Pyrococcus woesei) thermostable DNA polymerase. Alternatively the ProSTAR™ system from Stratagene™ may be used.

- 30 The PCR reaction is carried out in a PCR mixture that generally comprises the following: the template DNA (which will be amplified in the event of virus

-5-

detection), one or more primer pairs specific for members of a virus family, a thermostable polymerase enzyme (typically a DNA polymerase, such as Taq polymerase), deoxynucleotide triphosphates (dATP, dTTP, dCTP and dGTP) and a suitable buffer.

- 5 The PCR reaction generally comprises cycles of the following steps: a denaturation step, a primer annealing step and a polynucleotide synthesis step. Typically the PCR reaction comprises at least 25 cycles, such as 30, 35, 40 or more cycles, up to a maximum of 60 cycles for example. Generally, in the denaturation step, the PCR mixture is heated to a temperature at which the DNA in the PCR
- 10 mixture (in particular the region to be amplified) denatures to single-stranded form. The denaturing temperature is generally from 85 to 98°C.

- In one embodiment the PCR reaction comprises a "hot start" in which the PCR mixture is kept at the denaturing temperature for an extended amount of time before commencement of the thermal cycles, such as for 5 to 30 minutes, preferably
- 15 10 to 20 minutes. The use of Amplitaq Gold™ DNA polymerase (Applied Biosystems™) is preferred when the PCR reaction comprises a hot start.

- In the primer annealing step the primers bind to template nucleotide sequence in a sequence specific manner. This step is generally carried out at a temperature of from 30 to 65°C. In the polynucleotide synthesis step the polymerase replicates/
- 20 synthesises nucleotide sequence based on template sequence by addition of nucleotides to the 3' end of the bound primers. This step is generally carried out at about 72°C.

- In the method of the invention, the sample (generally after processing as described above) is subject to PCR conditions using a panel of multiple pairs of
- 25 degenerate primer pairs. In the course of a PCR reaction such primers are capable of binding the conserved sequences of the genome of a family of viruses. These conserved regions typically have a role in providing a necessary or advantageous activity or property to the virus. Generally, the conserved sequences may be coding or non-coding sequences.

- 30 In one embodiment the conserved sequences code for or are from virus proteins which have the following activities: DNA or RNA polymerase (replicase),

topoisomerase (helicase/gyrase), endonuclease (integrase), nucleic acid binding protein, protease, transcription factors, envelope glycoproteins, structural protein (e.g. capsid or nucleocapsid protein).

As discussed above multiple pairs of primers are used in the method each of the primer pairs used being selective (or specific) for members of a virus family (for example selective for a subfamily or genus). In the disclosure below regarding the numbers of primers used in different embodiments of the invention it is understood that this refers to the numbers of primers which are substantially specific for members of a virus family. However, in some embodiments additional primer pairs may be used which are selective for more than one family (for example selective for 2 to 10, such as 3 to 6 families). Such embodiments are within the scope of the present invention.

The panel of primer pairs may comprise sets of primer pairs which perform a nested PCR reaction. Generally such a set of primer pairs comprises a first and second primer pair. The first primer pair is able to amplify a template nucleotide sequence from a virus to form a PCR product. The second primer pair is able to amplify a nucleotide sequence using the PCR product generated by the first primer pair as a template. Multiply nested sets of primer pairs may also be used. The use of nested sets of primer pairs allows increased sensitivity and specificity.

The panel of primers used is capable of detecting viruses which are single-stranded or double-stranded DNA or single-stranded or double-stranded RNA viruses. The viruses are generally capable of infecting prokaryotic or eukaryotic cells, such as bacterial, animal, plant, yeast or fungal cells. Preferably the viruses are mammalian (preferably primate) or avian viruses, such as human, pig, horse, sheep, goat, cow, chicken, turkey or duck viruses.

The viruses are typically from any combination of the following families: Adenoviridae, Arenaviridae, Arteriviridae, Astroviridae, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Deltavirus, Filoviridae, Flaviviridae, Hepadnaviridae, Herpesviridae, Orthomyxoviridae, Papovaviridae, Paramyxoviridae, Parvoviridae, Picornaviridae, Polydnaviridae, Poxviridae, Reoviridae, Retroviridae, Rhabdoviridae, Togaviridae or Bornavirus.

-7-

Typically in the method 12 to 300 different primer pairs are used, such as 24 to 200 or 48 to 100 primer pairs. These primers may all be used in the same multi-well plate (placed on a thermal cycling machine). The plate may be a 96-well or 384-well plate.

- 5 In a preferred embodiment at least one of the wells in which the PCR is done comprises more than one primer pair, such as 2, 3, 4, 5, 6, 7, 8 or 9 primer pairs. Typically 3 to 96, such as 12 to 48, of the wells comprise more than one primer pair.

In one embodiment, some or all of the primer pairs used in the same well carry different labels. Thus, one or both primers of each primer pair carries a label.

- 10 When both primers of a primer pair carry a label then these labels are different from each other. Typically, at least one of the primers in each primer pair will carry a different label from that used for the other primer pairs in the same well. The PCR product generated by labelled primers carries the labels present on the primers.

- Thus, after different primer pairs have been used for PCR in the same well
15 detection of the labels in the PCR products can be used to deduce which primer pair has directed the PCR reaction. In one embodiment all forward primers of the group are labelled with one colour and the reverse primers are labelled with a different colour.

- In a preferred embodiment the primers are labelled with a fluorescent label,
20 such as fluorescein based labels (e.g. fluorescein isothiocyanate). Different primer pairs may be labelled with fluorescent labels of different colours. The fluorescent labels which are used may be capable of detection by a Beckman Coulter CEQ2000™ or Applied Biosystems A3700™ fluorescent DNA analyser. The fluorescent labels may be obtained from Beckman Coulter™ or Applied
25 Biosystems™.

- Another way of being able to determine which PCR products are generated by which primer pair is for each primer pair in the group to generate a PCR product of different size to the PCR products generated by the other primer pairs of the group. Typically each PCR product which is generated by the group of primers
30 differs in size from all the other PCR products by at least 20, such as at least 50, 100, 200, 500, 1000 or more nucleotides. Each PCR product may for example differ in

size from all other PCR products by up to a maximum of 3000 nucleotides.

In a preferred embodiment, multiple biological samples are screened simultaneously by subjecting DNA from multiple samples to PCR conditions using simultaneously multiple pairs of primers. Generally each of the samples is from a different (typically human) individual. Typically 2 to 80, such as 5 to 40 samples are screened simultaneously in the method.

In one embodiment, DNA from multiple samples is mixed together before being subject to PCR conditions. Typically 2 to 10 such as 5 to 8 samples are pooled together in this way.

After the DNA has been subject to PCR conditions any PCR product which is obtained may be sequenced. Typically prior to sequencing the PCR product is gel purified and cloned into a vector, for example a plasmid or a bacteriophage vector. Suitable plasmids are known and commercially available, such as pBluescriptTM (Stratagene) and pGEM-T-EasyTM (Promega). Suitable bacteriophage include bacteriophage λ and M13. Alternatively the sequencing reaction may be carried out on the PCR product itself, for example using one of the PCR primers as a sequencing primer.

Preferably an automated sequencer is used to obtain the sequence of the PCR product, such as a Beckman Coulter CEQ2000TM or Applied Biosystems A3700TM DNA analyser.

Designing the primers

Each of the primer pairs used in the method of the invention binds a sequence conserved across members of a virus family and selectively directs amplification of sequence from the members of the family. The multiple primer pairs which are used are typically designed by:

- (i) providing a plurality of amino acid sequences from members of a first virus family,
- (ii) comparing the sequences to identify conserved regions,
- (iii) designing a first primer pair using a computer based method, wherein each primer in the pair binds a nucleotide sequence that encodes a conserved region

identified in (ii) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the nucleotide sequences that encode conserved regions in members of the first virus family, and

(iv) repeating steps (i) to (iii) for each virus family.

5 The multiple primer pairs may also be designed by:

(i) providing a plurality of nucleotide sequences from members of a first virus family,

(ii) comparing the sequences to identify conserved regions,

(iii) designing a first primer pair using a computer based method, wherein

10 each primer in the pair binds a conserved region identified in (ii) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the conserved regions in members of the first virus family, and

(iv) repeating steps (i) to (iii) for each virus family.

The multiple pairs of primers are capable of detecting unknown viruses in a
15 sample, wherein such a sample originates from a single individual or is a pooled sample from individuals of the same species. Thus the panel of primers detects viruses which infect the same species.

The number of primers designed by the above steps is typically the same as the numbers of primers mentioned above for use in the method of the invention. The
20 primer pairs which are designed bind sequence which is conserved across members of a virus family. The panel of primer pairs which is designed may comprise primer pairs that bind sequence which is conserved across substantially members of the family or across a subset of the members of the family, for example across all members of a subfamily or of a genus.

25 Generally, the primer pairs bind at least 70%, at least 80%, or at least 90% of the known viruses of the family, subfamily or genus. Preferably less than 10, such as less than 5, primer pairs will be used for the detection of any given family, subfamily or genus in the panel.

The panel of primer pairs is generally capable of detecting viruses from at
30 least 10, 15, 20, 30 or more families, typically up to a maximum of 35 families.

The panel of primer pairs may comprise sets of primer pairs which perform a

nested PCR reaction. Generally such a set of primer pairs comprises a first and second primer pair. The first primer pair is able to amplify a template nucleotide sequence from a virus to form a PCR product. The second primer pair is able to amplify a nucleotide sequence using the PCR product generated by the first primer pair as a template. The use of nested sets of primer pairs allows increased sensitivity.

In a preferred embodiment each primer pair is specific for a particular virus family, so that it does not detect viruses of other families.

The plurality of amino acid or nucleotide sequences are provided from different known viruses of the same family. The sequences will be for the same protein of the different viruses. Typically at least 5, 10, 20, 50, 100 or more sequences are provided. The maximum number of sequences provided will, for example, be 300 sequences.

Each of the sequences which is provided is typically at least 20, 50, 100, 200 or more amino acids or nucleotides in length. In general the maximum length of the nucleotide sequences is 1000 nucleotides and the maximum length of the amino acid sequences is 300 amino acids. The sequences may be obtained from a database of sequences, such as GenBank. The sequences may be obtained from a database comprising virus sequences which are organised into homologous protein families (based on sequence similarity relationships).

In a preferred embodiment the sequences are obtained from the VIDA database (described in Alba et al (2001) Nucleic Acids Research 29, 133-136) or the Virus Division of GenBank. The sequences may be provided in the form of a database, preferably in computer-readable form. The sequences are preferably provided in the form of a computer-readable database constructed using programs which identify homologous protein families, such as GeneTableMaker, MKDOM or PSCBuilder.

The sequences which have been provided are compared to identify conserved regions. Typically such conserved regions will have a length of at least 12 nucleotides, such as at least 15, 21, 27, 36, 99 or more nucleotides (generally up to a maximum length of 200 nucleotides) or at least 4, 5, 7, 10, 25 or more amino acids (generally up to a maximum length of 50 amino acids).

Across the conserved region the virus sequences which are being provided will of course share identity or similarity. Typically the amino acids or nucleotides in at least 50% of the positions in the region will be the same in at least 50 %, 60%, 70%, or 80% of the viruses of the group (i.e. in the family, genus or subfamily).

5 The algorithm which identifies conserved regions generally uses a multiple sequence alignment method. The method may comprise (a) aligning all pairs of sequences separately to calculate a distance matrix giving the divergence of each pair of sequences, (b) calculating a guide tree from the distance matrix, and (c) aligning the sequences progressively according to the branching order in the guide tree.

10 A preferred algorithm for the aligning the conserved sequences is CLUSTALW as described in Thompson et al (1994) Nucleic Acids Research 22, 4673-80. Other algorithms that can be used for aligning sequences are MultAlin (Corpet (1988) Nucleic Acids Research 16, 10881-90) or Jalview (Clamp et al (1998) <http://barton.ebi.co.uk>). BLOCKS of conserved regions of amino acids may be
15 extracted from the multiple alignments, typically using the program Blocks Multiple Alignment Processor. Alternatively the entire process of performing multiple alignments and extracting BLOCKS can be performed using BLOCKMAKER (Henikoff and Henikoff (1994) Genomics 19, 97-107).

20 The output from the alignment and BLOCK extraction set (i.e. the information describing the identified conserved regions) is then entered into the algorithm which designs the primers. Such output is typically in the form of partial sequences which correspond to the conserved regions (BLOCKS). These BLOCKS are input into a primer design algorithm. In one embodiment such an algorithm is CODEHOP.

25 In the primer design step the conserved regions which are chosen as targets for primers preferably comprise few codons with degenerate counterparts, i.e. preferably the sequence has a low redundancy, such as a redundancy of less than 512 fold, 256 fold or 128 fold. Each primer binds in accordance with Watson-Crick base pairing and thus the binding is sequence specific. Each primer will thus be designed
30 to be wholly or partially complementary to the sequence to which it binds.

Each of the primers typically has a length of at least 8 nucleotides, such as at

least 10, 12, 15, 20, 30, 40 or more nucleotides (up to a maximum of 50 nucleotides for example). In one embodiment the primer may comprises at least 2, 4 or 6, up to a maximum of 10 for example, inosine bases. Inosine is able to bind to any of the four nucleotides and therefore use of inosine causes a reduction in effective redundancy.

- 5 Each primer pair will be designed so that the PCR product generally has a length of at least 20, such as at least 50, 100, 200, 500, 1000 or more nucleotides (and typically up to a maximum of 5×10^3 nucleotides long).

Each primer is preferably designed so that it anneals to a single site, i.e. the primer will not bind to any other site in the genome of the relevant viruses.

- 10 Each primer is preferably designed so that it does not exhibit secondary structure, i.e. the nucleotides in the primer will not bind substantially to any other nucleotide in the primer apart from those to which it is covalently linked. In addition preferably each primer is designed so that it does not bind other primers with the same sequence.

- 15 In one embodiment the 3' region, and preferably the 3' terminal nucleotide of the primer binds to the target sequence with high affinity, thus preferably this region or nucleotide comprises a G or C.

- Generally each primer is designed to have an annealing temperature of from 30 to 65°C, such as 50 to 60°C or 35 to 45°C. In addition each primer pair may be
20 designed to ensure that the two primers do not bind to each other.

The primers are designed by a computer based algorithm. In one embodiment such an algorithm designs primers according to the following rules:

- 1) A set of blocks is input, where a block is an aligned array of amino acid sequence segments without gaps that represents a highly conserved region of
25 homologous proteins. A weight is provided for each sequence segment, which can be increased to favour the contribution of selected sequences in designing the primer. A codon usage table is chosen for the target genome.
- 2) An amino acid position-specific scoring matrix (PSSM) is computed for each block using the odds ratio method.
- 30 3) A consensus amino acid residue is selected for each position of the block as the highest scoring amino acid in the matrix.

4) For each position of the block, the most common codon corresponding to the amino acid chosen in step 3 is selected utilizing the user-selected codon usage table. This selection is used for the default 5' consensus clamp in step 8.

5) A DNA PSSM is calculated from the amino acid matrix (step 2) and the
5 codon usage table. The DNA matrix has three positions for each position of the amino acid matrix. The score for each amino acid is divided among its codons in proportion to their relative weights from the codon usage table, and the scores for each of the four different nucleotides are combined in each DNA matrix position. Nucleotide positions are treated independently when the scores are combined. As an
10 option, the highest scoring nucleotide residue from each position can replace the most common codons from step 4 that are used in the consensus clamp.

6) The degeneracy is determined at each position of the DNA matrix based on the number of bases found there. As an option, a weight threshold can be specified such that bases that contribute less than a minimum weight are ignored in
15 determining degeneracy.

7) Possible degenerate core regions are identified by scanning the DNA matrix in the 3' to 5' direction. A core region must start on an invariant 3' nucleotide position, have length of 11 or 12 positions ending on a codon boundary, and have a maximum degeneracy of 128 (this is the default setting of CODEHOP). The
20 degeneracy of a region is the product of the number of possible bases in each position.

8) Candidate degenerate core regions are extended by addition of a 5' consensus clamp from step 4 or 5. The length of the clamp is controlled by a melting point temperature calculation (the CODEHOP default is 60°C) and is usually about
25 20 nucleotides.

9) Steps 7 and 8 are repeated on the reverse complement of the DNA matrix from step 5 for primers corresponding to the opposite DNA strand.

In one embodiment CODEHOP (Rose et al (1998) Nucleic Acids Research
26, 1628-1635) is used to design the primer pairs. This program uses the above
30 rules.

The primers designed by the algorithm may then be mapped back to the

-14-

original sequence to choose primer pairs which provide the desired length of PCR product.

The above-described computer based method is repeated until the desired number of primer pairs have been designed. Optionally the primer pairs can then be synthesised and tested. They are typically tested to determine the optimal conditions for using the primers in a PCR reaction.

In one embodiment the primers are tested for their ability to amplify one or more of the plurality of nucleotide sequences from known viruses which were used to design the primers, or in the case of amino acid sequences from known viruses being used to design the primers the primers may be tested for their ability to amplify the nucleotide sequence from the virus which encodes the amino acid sequence.

The primers may be tested in a range of buffer conditions to determine optimal buffer conditions for PCR using the primers. The buffer conditions which may be tested include pH (typically between 7 and 10), magnesium concentration (typically from 0.5 mM to 5 mM), potassium chloride (typically from 0 to 100 mM), ammonium chloride (typically 0 to 100 mM), glycerol (typically 0 to 20%), dimethylsulphoxide (typically 0 to 20%), ethanol (typically 0 to 20%), sorbitol (typically 0 to 20%) or betaine (typically 1M betaine).

The primers may be tested at a range of different temperatures to determine the optimal temperatures in the PCR reaction. Preferably the primers are tested in PCR reaction in which a range of primer annealing temperatures are tested. Typically the range of temperatures is from 30 to 65°C.

The panel of primer pairs or a group of primers within the panel may be designed to be used together on the same plate (i.e. using the same thermal cycles). Thus such primer pairs will be designed to work at the same annealing temperature.

In one embodiment a group of primer pairs within the panel are designed to have similar optimal conditions for use in PCR so that they can be used optimally in the same well or reaction vessel, i.e. that they can be used in multiplex PCR. Such a group typically comprises at least 2, 3, 4, 5, 6 or more primer pairs (up to a maximum of 8 primer pairs for example).

To provide such primer pairs the computer based method steps may be used

to design primer pairs which are calculated to have similar annealing temperatures and/or the primers are tested to select primer pairs which can be used optimally together. Such testing typically determines whether the primers work optimally with the same buffers and/or whether the primers have similar annealing temperatures.

5

Validating a PCR product as being from a novel virus

After sequencing the PCR product(s), the next step is to determine whether each sequence is present in at least one database of known nucleic acid sequences, typically sequences of viruses known to infect the individuals from which the samples are derived. Appropriate databases include the virus subdivision of GenBank or the VIDA database.

In addition each sequence is typically also compared with a database of human sequences to exclude sequences which are human sequences. Such a database is generally a comprehensive or consensus human genome database. Preferably, at least one of the human sequence databases searched contains an essentially complete human genome sequence. However, it needs to be borne in mind that, although there has recently been a great deal of publicity about the "completion" of the human genome sequence, not all the human genome has in fact been sequenced, and it is possible that a cloned sequence could fall within the unsequenced part of the genome. The human genome contains large areas with repetitive sequences, and much of the unsequenced genome is within these areas.

In order to make as comprehensive a search as possible, it is desirable to search a range of different types of database; in addition to a human genome database, it is desirable to search, for example, a database comprising expressed sequence tags (ESTs) and a database comprising repetitive elements of the human genome. Appropriate databases include GenBank, the EMBL database, the Celera human genome database, the Ensemble human genome database, the DNA Data Bank of Japan (DDBJ), the Incyte LifeSeq™ database of ESTs and the Repbase database of repetitive elements in the human genome.

Where the sequence is found to be not present in any of the interrogated databases of known sequences, this indicates that the nucleic acid may be from a

-16-

previously unknown virus. The nucleic acid then becomes a candidate for further investigation and may be designated a Primary Candidate Virus (PCV).

It is generally necessary to confirm by experimentation that a nucleic acid sequence designated a PCV is not in fact a human sequence. A preferred way of
5 doing this involves designing and synthesising a specific primer set (or sets) to amplify the nucleic acid designated a PCV and determining whether the set(s) are able to amplify any DNA in a sample of complete genomic human DNA. The amplification conditions for each primer set may be optimised using the original sample from which the PCV derives or using the PCR product which is obtained in
10 the method of the invention.

The primer set may be used to screen one or more samples of human genomic DNA, for example from 1 to 100 samples, preferably from 5 to 50 samples. As an alternative to PCR, human genomic DNA may be probed with a labelled probe containing sequence from the original PCR product (e.g by Southern blotting).

15 If the PCV cannot be detected in human DNA by experimentation (by PCR or hybridisation with a labelled probe), it may then be subjected to further analysis. It may be designated a Secondary Candidate Virus (SCV).

The further analysis of an SCV may include gene walking to determine whether the original cloned nucleic acid sequence exists in nature as part of a longer
20 sequence, such as the genomic sequence of an unknown virus. Gene walking may be carried out using techniques known in the art, such as vectorette PCR (Allen et al, PCR Methods Appl. 4:71-75), rapid amplification of cDNA ends (RACE, Frohman et al Proc Natl Acad Sci U S A. 85:8998-9002), rapid amplification of genomic ends (RAGE, Cormack and Somssich. 1997. Gene. 194:273-276) and methods derived
25 from these. Alternatively, the SCV sequence may be "extended" by screening a DNA or cDNA library using the original cloned nucleic acid sequence as a probe.

The additional sequence information obtained through DNA walking may reveal information about the identity of the SCV which cannot be determined from the original clone. The additional information may therefore be analysed, for
30 example to determine whether it contains an open reading frame (i.e. a sequence encoding a protein); the presence of an open reading frame provides further support

-17-

for the suggestion that the SCV is a virus. Furthermore, the additional information may identify the SCV as being related to a known virus; for example, the information may identify the SCV as being a new member of a known family of viruses.

- A further step may then be to determine whether a newly-identified candidate virus is associated with a disease, for example with a cancer, autoimmune disease, cardiovascular disease or other disease mentioned above. This may be done by obtaining a specimen from each member of a group of subjects with a disease; determining whether the cloned nucleic acid or other nucleic acid of the same virus is present in each specimen; and determining whether the proportion of subjects in whom the nucleic acid is present is greater in the group of subjects who have the disease than in a control group of subjects who do not have the disease, wherein a said greater proportion suggests that the virus may cause or contribute to the disease.

- Typically, the process of determining whether the nucleic acid is present or absent from a specimen from a subject may be carried out by PCR using primers specific for the novel sequence (including any contiguous sequence obtained by DNA walking). Initially, perhaps from 10 to 50 patients from a disease group may be tested, but if positive results are obtained in initial studies, the investigation may be extended to a larger group (e.g. a group of up to 100, 500, 1000 or 10,000). The nature of the biological specimens taken from the members of the group varies depending on the disease association that is being investigated; where possible specimens are from disease affected tissue and from peripheral blood of the subjects (for a published example of this see Griffiths et al, 1999, Arthritis Rheumatism, 42:448-454). The specimens may be from the same tissue and fluid types as the biological samples used in the initial screening assay described above.

- Once a new virus has been identified and found to be positively associated with a particular disease or condition, serological and genetically-based diagnostic assays for infection by the virus may readily be devised. Genetically-based assays can be developed by using the nucleotide sequence of the virus to design probes and/or PCR primers for specifically detecting the nucleic acid of the virus.
- Serological assays can be developed by producing recombinant proteins or protein fragments encoded by the virus and testing for the presence of antibodies to these

-18-

proteins in human sera. Alternatively, antibodies specific for the proteins of the virus may be made and the antibodies used to detect the virus directly. The serological assays may take the format of an ELISA, western blot or immunofluorescence assay. Correlations may be sought between serological data and genetic data. Furthermore, the organism provides a target for the development of therapies and/or prophylactic vaccines against the disease.

The following Example illustrates the invention:

Example

The Example below refers to Figure 1 which shows how primers were designed using a database known as 'VIDA', and computer programs known as 'CLUSTALW', 'BLOCKS' and 'CODEHOP'.

Designing a panel of primers

A panel of primers was designed for detecting unknown viruses from the family Herpesviridae according to the strategy shown in Figure 1. The amino acid sequences of herpes virus DNA packaging protein UL15 were obtained from the VIDA database (Alba et al, see above). These sequences are shown in Table 1.

The sequences obtained from the VIDA database were then imported into CLUSTALW. This compares the protein sequences to identify conserved regions and then aligns the sequences according to the conserved regions. The alignment produced by CLUSTALW is shown in Table 2.

The BLOCKS program was then used to extract the sequences of the conserved regions identified by CLUSTALW, and to enter these sequences into CODEHOP. The primer sequences were then designed by CODEHOP using the conserved sequences. The output from the CODEHOP program is shown in Table 3.

Table 1. All protein sequences of DNA packaging protein UL15 extracted from VIDA.
Here written as a list and unaligned.

>gi_10180719
5 MFGGLLGEETKRFHERLIMKTKNDRLGASHRNERSIRDGMVDAPFLNFAIVPRRHQTVMPAIGILHNCC
DSLGLYSAITTMYLYSSIAACEFDELREDSVERCYPRITNAQAFLSPMMRVANSIIIPQYEDMECAHR
NAYYSTNNSFISMRTSDAFKQLTVFISRFSKLLIASFRDVNKLDDHTVKKRARIADAPSYKLGHTLELPQ
XMLMHTATYFVTSVLLGDHAERASERLLRVADFTPHFSDIVTRHFRQRAATVFLVPRRHGKTWFLVPLIALA
MSSFGIRIGYVTSIRKALIEPVFEDIQDLRRWFGAHRVDHVKGETITFSFPGLESTVTFASSHNTNSI
10 RQQDNLLFVDEANFIRPDAVQTIIGFLNQATCKIIFVSSTNSGKASTSFLYGLKGSADOLLNVVTYICD
EHMKIVDTYTNATSCSCVYLNKPVFTMDGMRRTAEMFLPDSFMKSIIGGVVDRTICQGRSIFTASA
IDRFLIYRSTVNMQDFFSQDLVYVVDPAFTANTKASGTGVAVIGKYGTDIYVGLHEHYFLAALTGESSD
SIGYCVACQLIQICAIHERKRFQVICKIAIEGNSNQDSAVAIAIRIAIEMIYNKAAVAFTPHNVSYFYSKS
NGTDVEYFFLLQRQKTAADFDFIAQFNSGRVLASQDLVSTTVSLTDPVEYLTKQLTNISEVVTGPTCT
15 RTFSGKKGNDOTVVALTMVVISAHIPMAFAPIRV
>gi_7673189
MFGGLLGEETKRFHERLIMKTKNDRLGASHRNERSIRDGMVDAPFLNFAIVPRRHQTVMPAIGILHNCC
DSLGLYSAITTMYLYSSIAACEFDELREDSVERCYPRITNAQAFLSPMMRVANSIIIPQYEDMECAHR
NAYYSTNNSFISMRTSDAFKQLTVFISRFSKLLIASFRDVNKLDDHTVKKRARIADAPSYKLGHTLELPQ
20 XMLMHTATYFVTSVLLGDHAERASERLLRVADFTPHFSDIVTRHFRQRAATVFLVPRRHGKTWFLVPLIALA
MSSFGIRIGYVTSIRKALIEPVFEDIQDLRRWFGAHRVDHVKGETITFSFPGLESTVTFASSHNTNSI
RQQDNLLFVDEANFIRPDAVQTIIGFLNQATCKIIFVSSTNSGKASTSFLYGLKGSADOLLNVVTYICD
EHMKIVDTYTNATSCSCVYLNKPVFTMDGMRRTAEMFLPDSFMKSIIGGVVDRTICQGRSIFTASA
IDRFLIYRSTVNMQDFFSQDLVYVVDPAFTANTKASGTGVAVIGKYGTDIYVGLHEHYFLAALTGESSG
25 SIGYCVACQLIQICAIHERKRFQVICKIAIEGNSNQDSAVAIAIRIAIEMIYNKAAVAFTPHNVSYFYSKS
NGTDVEYFFLLQRQKTAADFDFIAQFNSGRVLASQDLVSTTVSLTDPVEYLTKQLTNISEVVTGPTCT
RTFSGKKGNDOTVVALTMVVISAHIPMAFAPIRV
>gi_5669285
MFGGLGESAKKHFERLLDRNERLGASRNECLAROGSLVDAFLNFAISVPRRHQTVMPAVGTLHDCC
30 DGTGYSIAIATRLIYAGIVSEFGEVRRESLNGHISKRNREALLAFTLTKVANSITFHEYDQAQCAHR
NAYYSTNNSFISMRTSDAFQQLASFIIDRFSKLLASFKDVLNLDENNAKKRARIAPSYKLGHTLELPQ
XMLMHTATYFVTSVLLGDHAERASERLLRVADFTPHFSDIAATRHFRQRAATVFLVPRRHGKTWFLVPLIALA
MSSFGIRIGYVTSIRKALIEPVFEDIQDLRRWFGAHRVDHVKGETITFSFPGLESTVTFASSHNTNSI
RQQDNLLFVDEANFIRPDAVQTIIGFLNQATCKIIFVSSTNSGKASTSFLYGLKGSADOLLNVVTYICD
35 EHMKIVDTYTNATSCSCVYLNKPVFTMDGMRRTAEMFLPDSFMKSIIGGVVDRTICQGRSIFTASA
VERLLIYRSTVNMQDILSRDLVYVVDPAFTANTKASGTGIAVIGRYGADYIIVGLHEHYFLAALTGESSAD
AIGCEAAQCIQICAIHERKRFQVIRVAVGNSNQDSAVAIAIRISIDLASVYQSGVAPAFHDVCYFYSK
AGSNVEYFFLLQRQKTAADFDFIARFNSGRVLASQDLVSTTVSLTDPVEYLTKQLTNISEVVTGPTCT
40 RTFSGKKGNDOTVVALTMVVISAHISADATFAPIRGVEATCRGPTEA
>gi_1869837
MFGQQLASDVQVYLERLEKQRQKQGVGVDASAGHITGGDALRVPLFDATATPKRHQTVVPGVGTLDHCC
EHSPLFSAVARRLLFNSLVPAQLGRDQFGDHTAKLEFAPLAVARAVLAFRECAPEDAVPQRNAYSV
LNTFQALRESNAPQLVHFPRDQALLKTSFRASSLAETGPEKKRAKVDVATHTGQYGTLELQPMXIM
45 HATYFLAAVLLGDHAEQVNTFLRLVPEIPLFSDTAVRHFRQRAATVFLVPRRHGKTWFLVPLIALASLAS
GIKIQYTAIRIKATESPVFEDIACLRGWPQSSRDHVKGETISFSPFGDSRSTVTFASSHNTNSIIRQDF

-20-

NLLFVDEANFIRPDAVQTIMGFLAQANCKII FVSSNTNGKASTSFLYNLRGADELINAVVTYICDEHMPR
 VVTHTNATACSCYILNKPVFTIMDGAVERRTALFLPDSFMQEIIGGQARETQDRPVLTKSAGEREFLLYR
 PSTTTNSGLNAFELYVYVDPAFTANTRASGTGIAVVGRYRDDYII FFALEHFFLALTGSAADIARCVVH
 SLQVLALHPGAFRVRVAVBGNSSQSDAAVAIATNVTHTMHRLASAGANGPGPELLFTHCEPFGSVA
 5 PFFLLNKQKTAFEFHFIKKFNSGGVMSQELVSVTVRLQTDPEVYLLQNLNLTETVSPTNDVRYSGKR
 RNASDDMLVALVMAVMAIPLAAPTGI PPAFFITRTS
 >gi_59501
 MFGQQLASDVQYLERLEKQRLKVGAEASGLTMGGDALRVFFLD FATATFKRHQTVVPGVGTLDCC
 EHSPLFSAVARILLFSLVPAQLKGRDFGCDHTAKLEFLA PELVRAVARLEFKCAPADVVPORNAVYSV
 10 LNTFQALHRSBAFRQLVHFVRDFAQLLKTSSFRASSIETTGPPKKRAKVDVATHGRTYGTLELPQSMILM
 HATYFLAAVLLGDHABQVNTFLRLVFEIPLFSDAAVRHFRQRTATVFLVPRRHQKTWFLVPLIALSLASFA
 GIKIGYTAHIRKATEPVFEEIDARLGRWFGSARVDHVKGETISFSPDGSRTIVFASSHNTINGRQDP
 NLLFVDEANFIRPDAVQTIMGFLAQANCKII FVSSNTNGKASTSFLYNLRGADELINAVVTYICDEHMPR
 VVTHTNATACSCYILNKPVFTIMDGAVERRTALFLPDSFMQEIIGGQARETQDRPVLTKSAGEREFLLYR
 15 PSTTTNSGLNAFELYVYVDPAFTANTRASGTGIAVVGRYRDDYII FFALEHFFLALTGSAADIARCVVH
 SLQVLALHPGAFRVRVAVBGNSSQSDAAVAIATNVTHTMHRLASAGADAGEGPELLFTHCEPFGSVA
 YPFFLLNKQKTAFEFHFIKKFNSGGVMSQELVSVATVRLQTDPEVYLLQNLNLTETVSPTNDVRYSGKR
 RNASDDMLVALVMAVMAIPLAAGP PHTFAPITRVS
 >gi_2605992
 MFGKALSRETIQYFETLRKEVQSRSGAKNRRAEAQTGGEDDVKTAFINFAIPTPQRHQTVPVPGVGTLDCC
 CETAGIFASVARILLFSLKWNRGESKERLDPSSVEAYVDPKVEQALKTISFVEYNDAEARSCHNAVYS
 IMNTFDALRSSDAFHQVANFVARFSLVDTSPNGADLDGQQQTSKRILVDVPTYGKRGTELEFQNMIL
 MHATYFIAAVILGDHADRIGAFILNVFNTPEPSDATIRHFQRATVFLVPRRHQKTWFLVPLIALALATF
 20 KGIKIGYTAHIRKATEPVFDEIGARLQWFGNSFVDHVKGENTISFSPDGSKSTIVFASSHNTINGRQDP
 NLLFVDEANFIRPEAVQTTIIGFLNQTNCKII FVSSNTNGKASTSFLYNLRGADELINAVVTYICDEHME
 RVKATNTATSCSCYILNKPVFTIMDGAVERRTALFLPDSFMQEIIGGQHSIQAHRDEPVFTKTAQDRFL
 YRPSVAVARQDIMSBNLYVYVDPAFTTNMASGTGVAVVGRYRSHNVFGLHEFFLSALTGSSABLIARCV
 AQCLAQVFAIHSRPFDSVRVAVBGNSSQDAVAIATNIQLELNTLRQADVVMHPTGVLFHCTPPGSSVA
 25 YPFFLLQKQKTAFDFHFIKAFNSGLVLASQELISHTVRLQTDPEVYLLQNMNLTETVITGETSETRVFTGK
 RNASDDMLVALVMAVMAIPLPTTNAFSSLSQ
 >gi_330792
 MFGRVLGRETQYFEALRREVQARRGAKNRRAEAQNGEDDAKTAFINFAIPTPQRHQTVPVPGVGTLDCC
 CETAGIFASVARILLFSLKWNQGEARERLDPASVEAYVDPKVRQALKTISFVEYSDDEARSCHNAVYS
 IMNTFDALRSSDAFHQVASFVARFSLVDTSPNGADLDGQQQASKRARVDVPTYGKRGTELEFQNMIL
 30 MHATYFIAAVILGDHADRIGAFILNVFNTPEPSDATIRHFQRATVFLVPRRHQKTWFLVPLIALALATF
 KGIKIGYTAHIRKATEPVFDEIGARLQWFGNSFVDHVKGENTISFSPDGSKSTIVFASSHNTINGRQDP
 NLLFVDEANFIRPEAVQTTIIGFLNQTNCKII FVSSNTNGKASTSFLYNLRGAADDLINAVVTYICDEHME
 RVKATNTATSCSCYILNKPVFTIMDGAVERRTALFLPDSFMQEIIGGGRVSGAHRDEPVFTKTAQDRFL
 YRPSVAVARQDIMSBDLYVYVDPAFTTNMASGTGVAVVGRYRENWVFGMEHFFLSALTGSSABLIARCV
 35 AQCLAQVFAIHSRPFDSVRVAVBGNSSQDAVAIATNIQLELNTLRQADVVMHPTGVLFHCTPPGSSVA
 YPFFLLQKQKTAFDFHFIKAFNSGLVLASQELISHTVRLQTDPEVYLLQNMNLTETVITGETSETRVFTGK
 RNASDDMLVALVMAVYLSSLPFTSDAFSSLPQ
 >gi_971317
 MFGQAVGEQSARVFCRLLERRQRRAAERGARPDGQGARGEDDARVFFLDFAVAAPKRHQTVPVPGVGTLDH
 GYCELAFLAATASRLLLTSMARAEAGLNTGTGEAHSRETAGVLSALFPAHHPPAERAARCAHYSVMA
 40 ALESHRASGAFAGAAVAFVARFORLVSTSTFSLGGQDDADPRAKARVPEPSQQTGRGALEFQNMILPA

-21-

- TYFVAATLLGEHAERIGAPLRVA¹NTPDFSDAAVAHFRQ²ATVFLVPRRHGKTWFLVPLIALALATFSGI
KIGVTAHIRKATEPVFEEI³VARLRQMFGEERVDHVEGEVIS⁴FPDGA⁵RSITV⁶FASSEN⁷INGIRQ⁸QDN⁹LL
LFDVEMFIRPEAVQTI¹⁰VGLFQASCKI¹¹IFVSS¹²TNTGKAST¹³FLYNLKGASDGLNVVTYI¹⁴CNEHSTPRVA
AHGGTATACS¹⁵CYVLNKPVFT¹⁶ITMDAA¹⁷RN¹⁸TAE¹⁹TFL²⁰NSFM²¹QEI²²IGOC²³VARRAEPAAV²⁴TRANGE²⁵QFLLVRP
STAAR²⁶GF²⁷WFERLYMID²⁸PAFTSNARASGSGIAV²⁹VGRHRS³⁰SWLV³¹GLEHFFL³²PALTGS³³SAAS³⁴LARC³⁵AVRC
FAQVMA³⁶VERREL³⁷DGLFVA³⁸VEGNSQDSAVAIAL³⁹GVRE⁴⁰LD⁴¹SLAAGAV⁴²MPAST⁴³FPYHCR⁴⁴PPGSAV⁴⁵VP
FL⁴⁶LQKQ⁴⁷KTAA⁴⁸FDH⁴⁹FIRL⁵⁰NSGR⁵¹VVASQ⁵²DLAS⁵³LTVRL⁵⁴QTD⁵⁵PVEYLF⁵⁶EQ⁵⁷LQNL⁵⁸TESTAG⁵⁹PGGARAF⁶⁰SGK⁶¹RG
AADD⁶²LMVALVMA⁶³VFV⁶⁴SGL⁶⁵PPTD⁶⁶GA⁶⁷CF⁶⁸LA⁶⁹FR⁷⁰PAD
>gi_5869808
10 MSLIMFGR⁷¹TLG⁷²ESV⁷³RY⁷⁴PERL⁷⁵KRR⁷⁶DER⁷⁷FT⁷⁸LES⁷⁹PT⁸⁰PCSTRQ⁸¹SG⁸²BNAT⁸³QI⁸⁴FF⁸⁵IN⁸⁶FAID⁸⁷VT⁸⁸RRHQA⁸⁹VI⁹⁰PG
IG⁹¹TLEN⁹²CEY⁹³LPL⁹⁴FSAT⁹⁵ARRAM⁹⁶GF⁹⁷AL⁹⁸ST⁹⁹GV¹⁰⁰NT¹⁰¹ENV¹⁰²LK¹⁰³WRY¹⁰⁴SV¹⁰⁵NAN¹⁰⁶VS¹⁰⁷PEL¹⁰⁸KA¹⁰⁹VSS¹¹⁰Q¹¹¹VEY¹¹²SE
ZAA¹¹³PHRD¹¹⁴YSG¹¹⁵WM¹¹⁶TF¹¹⁷FR¹¹⁸AS¹¹⁹LS¹²⁰DS¹²¹FC¹²²QL¹²³ST¹²⁴FT¹²⁵QR¹²⁶PSY¹²⁷L¹²⁸VETS¹²⁹PES¹³⁰IECC¹³¹SG¹³²HG¹³³KRA¹³⁴VD¹³⁵VPY¹³⁶GRY¹³⁷KG
EL¹³⁸QPK¹³⁹ML¹⁴⁰MHT¹⁴¹HF¹⁴²ISS¹⁴³VLL¹⁴⁴GD¹⁴⁵HADR¹⁴⁶VDC¹⁴⁷FLET¹⁴⁸VNT¹⁴⁹PT¹⁵⁰SV¹⁵¹SD¹⁵²SVLE¹⁵³HF¹⁵⁴KOK¹⁵⁵ST¹⁵⁶VEL¹⁵⁷VPRR¹⁵⁸OK¹⁵⁹TW¹⁶⁰FL¹⁶¹VPL
IAL¹⁶²VNAT¹⁶³FRG¹⁶⁴IKV¹⁶⁵QTAH¹⁶⁶IRKATE¹⁶⁷PV¹⁶⁸EQ¹⁶⁹IK¹⁷⁰SELE¹⁷¹QMF¹⁷²GANY¹⁷³VD¹⁷⁴HVGES¹⁷⁵IT¹⁷⁶FS¹⁷⁷TQ¹⁷⁸SY¹⁷⁹TA¹⁸⁰FASSEN
15 TNGIR¹⁸¹GQ¹⁸²DNLL¹⁸³FVDEAN¹⁸⁴FIR¹⁸⁵PD¹⁸⁶AVQ¹⁸⁷TVG¹⁸⁸FL¹⁸⁹QNTCK¹⁹⁰II¹⁹¹FVSS¹⁹²TNTGKAST¹⁹³FLYNL¹⁹⁴RQSSD¹⁹⁵QLLNV¹⁹⁶VT
YV¹⁹⁷CD¹⁹⁸DM¹⁹⁹PR²⁰⁰LJES²⁰¹DV²⁰²TACS²⁰³CYVLNK²⁰⁴VP²⁰⁵ITMD²⁰⁶GA²⁰⁷MR²⁰⁸TAD²⁰⁹LPMADS²¹⁰FVQEI²¹¹VGR²¹²KQNS²¹³GGV²¹⁴GF²¹⁵PL²¹⁶FL
TK²¹⁷TARE²¹⁸RFIL²¹⁹RPST²²⁰VAN²²¹CAIL²²²SS²²³VL²²⁴YV²²⁵VD²²⁶PAFT²²⁷SN²²⁸TRAS²²⁹GT²³⁰OVAI²³¹VGRY²³²KSDWI²³³I²³⁴QGLE²³⁵HF²³⁶PLAAL²³⁷Q
TS²³⁸SEI²³⁹GR²⁴⁰CV²⁴¹QCL²⁴²GH²⁴³LAL²⁴⁴HT²⁴⁵FT²⁴⁶IN²⁴⁷VHS²⁴⁸IEGNS²⁴⁹QDSAVAIS²⁵⁰LA²⁵¹IA²⁵²Q²⁵³FA²⁵⁴VLEK²⁵⁵GR²⁵⁶NVL²⁵⁷SSA²⁵⁸PVLL²⁵⁹FY
HS²⁶⁰IP²⁶¹QCS²⁶²VAY²⁶³FF²⁶⁴LQ²⁶⁵KQ²⁶⁶KT²⁶⁷PAVD²⁶⁸YF²⁶⁹VK²⁷⁰RFNS²⁷¹GN²⁷²IAS²⁷³QEL²⁷⁴VS²⁷⁵LT²⁷⁶VKL²⁷⁷GV²⁷⁸DP²⁷⁹VEY²⁸⁰LK²⁸¹QCL²⁸²NL²⁸³ATE²⁸⁴VIK
20 QM²⁸⁵NLD²⁸⁶TK²⁸⁷YT²⁸⁸GRG²⁸⁹TT²⁹⁰QMSD²⁹¹DL²⁹²VAL²⁹³IMS²⁹⁴V²⁹⁵IGSS²⁹⁶CI²⁹⁷PD²⁹⁸SV²⁹⁹MP³⁰⁰IK
>gi_5708110
ML³⁰¹GKES³⁰²VEI³⁰³KY³⁰⁴RDAL³⁰⁵K³⁰⁶R³⁰⁷TH³⁰⁸ERG³⁰⁹PD³¹⁰VD³¹¹Q³¹²ENS³¹³DS³¹⁴N³¹⁵FIT³¹⁶TAS³¹⁷IC³¹⁸DN³¹⁹DS³²⁰ARD³²¹TNS³²²PAS³²³RFQ³²⁴FAID³²⁵VP
QR³²⁶QACIA³²⁷PI³²⁸GS³²⁹FN³³⁰CCAI³³¹SRA³³²PSY³³³MA³³⁴SEI³³⁵IYEN³³⁶LASY³³⁷TRY³³⁸TD³³⁹AL³⁴⁰NL³⁴¹Q³⁴²VS³⁴³PR³⁴⁴QL³⁴⁵FT³⁴⁶GA³⁴⁷DS³⁴⁸IL³⁴⁹S
PAL³⁵⁰RQ³⁵¹KL³⁵²ANL³⁵³MFAR³⁵⁴PAS³⁵⁵DSL³⁵⁶HD³⁵⁷KAFD³⁵⁸QIM³⁵⁹GV³⁶⁰RF³⁶¹VS³⁶²DE³⁶³FS³⁶⁴QL³⁶⁵AN³⁶⁶FT³⁶⁷YR³⁶⁸HT³⁶⁹LL³⁷⁰KK³⁷¹FS³⁷²GG³⁷³QAS³⁷⁴NDY³⁷⁵K
25 RAK³⁷⁶LEK³⁷⁷TT³⁷⁸SEQ³⁷⁹RD³⁸⁰TL³⁸¹EL³⁸²FQ³⁸³K³⁸⁴ML³⁸⁵MEAT³⁸⁶YFAS³⁸⁷IC³⁸⁸LG³⁸⁹EST³⁹⁰ERS³⁹¹NY³⁹²LS³⁹³TV³⁹⁴MT³⁹⁵SL³⁹⁶SE³⁹⁷NI³⁹⁸I³⁹⁹QH⁴⁰⁰FRQ⁴⁰¹RT
VFL⁴⁰²VPRR⁴⁰³HK⁴⁰⁴TW⁴⁰⁵FL⁴⁰⁶VPL⁴⁰⁷IS⁴⁰⁸LL⁴⁰⁹VSS⁴¹⁰FE⁴¹¹QIR⁴¹²GY⁴¹³TAH⁴¹⁴IRKATE⁴¹⁵PV⁴¹⁶IE⁴¹⁷PT⁴¹⁸LY⁴¹⁹MF⁴²⁰GA⁴²¹KQ⁴²²VE⁴²³Q⁴²⁴GET⁴²⁵IT⁴²⁶FE
TF⁴²⁷RN⁴²⁸KSA⁴²⁹IV⁴³⁰FAS⁴³¹QNT⁴³²NG⁴³³LQ⁴³⁴QDN⁴³⁵FL⁴³⁶VDEAN⁴³⁷FI⁴³⁸K⁴³⁹PAAL⁴⁴⁰HT⁴⁴¹VM⁴⁴²FL⁴⁴³NT⁴⁴⁴CK⁴⁴⁵LF⁴⁴⁶VSS⁴⁴⁷TNT⁴⁴⁸CS⁴⁴⁹HTS
L⁴⁵⁰YNL⁴⁵¹KOK⁴⁵²TNS⁴⁵³LN⁴⁵⁴VVTYI⁴⁵⁵CDE⁴⁵⁶H⁴⁵⁷MEI⁴⁵⁸Q⁴⁵⁹RTD⁴⁶⁰V⁴⁶¹TTCS⁴⁶²CYVL⁴⁶³HK⁴⁶⁴PV⁴⁶⁵FS⁴⁶⁶HD⁴⁶⁷SE⁴⁶⁸VNT⁴⁶⁹AD⁴⁷⁰L⁴⁷¹V⁴⁷²KDS⁴⁷³FM⁴⁷⁴HEIA
Q⁴⁷⁵RGAK⁴⁷⁶YDS⁴⁷⁷RL⁴⁷⁸TV⁴⁷⁹RAL⁴⁸⁰D⁴⁸¹Q⁴⁸²FLY⁴⁸³RPST⁴⁸⁴SS⁴⁸⁵KPH⁴⁸⁶IS⁴⁸⁷QL⁴⁸⁸KIL⁴⁸⁹TV⁴⁹⁰VD⁴⁹¹PAFT⁴⁹²TR⁴⁹³RSAS⁴⁹⁴GT⁴⁹⁵GI⁴⁹⁶AL⁴⁹⁷TL⁴⁹⁸EDS
30 MY⁴⁹⁹MGAE⁵⁰⁰HF⁵⁰¹YLDAL⁵⁰²TGEA⁵⁰³L⁵⁰⁴ETIA⁵⁰⁵Q⁵⁰⁶CVY⁵⁰⁷L⁵⁰⁸CIAY⁵⁰⁹CC⁵¹⁰LI⁵¹¹HAGA⁵¹²FRE⁵¹³IRIA⁵¹⁴VEGNS⁵¹⁵QDSAAAI⁵¹⁶AG⁵¹⁷NL⁵¹⁸TE⁵¹⁹LLDS
L⁵²⁰RR⁵²¹L⁵²²GF⁵²³SL⁵²⁴TF⁵²⁵ARS⁵²⁶Q⁵²⁷POT⁵²⁸AM⁵²⁹HP⁵³⁰FL⁵³¹LN⁵³²KQ⁵³³SA⁵³⁴FD⁵³⁵LV⁵³⁶SL⁵³⁷FN⁵³⁸SG⁵³⁹RF⁵⁴⁰AS⁵⁴¹QEL⁵⁴²VN⁵⁴³TL⁵⁴⁴VL⁵⁴⁵SK⁵⁴⁶DP⁵⁴⁷CEY⁵⁴⁸LV
DQ⁵⁴⁹IR⁵⁵⁰NT⁵⁵¹V⁵⁵²TH⁵⁵³Q⁵⁵⁴Q⁵⁵⁵PD⁵⁵⁶S⁵⁵⁷FR⁵⁵⁸TF⁵⁵⁹SG⁵⁶⁰K⁵⁶¹RV⁵⁶²PD⁵⁶³ML⁵⁶⁴VA⁵⁶⁵VM⁵⁶⁶STY⁵⁶⁷LA⁵⁶⁸EG⁵⁶⁹SP⁵⁷⁰AG⁵⁷¹HP⁵⁷²IA⁵⁷³PI⁵⁷⁴GR⁵⁷⁵Q⁵⁷⁶RP⁵⁷⁷A
>gi_1813970
ML⁵⁷⁸RGDSAA⁵⁷⁹KI⁵⁸⁰QERY⁵⁸¹AE⁵⁸²LQ⁵⁸³K⁵⁸⁴RK⁵⁸⁵SH⁵⁸⁶PTS⁵⁸⁷CIST⁵⁸⁸APT⁵⁸⁹N⁵⁹⁰VAL⁵⁹¹CR⁵⁹²KRY⁵⁹³Q⁵⁹⁴MM⁵⁹⁵FL⁵⁹⁶GLA⁵⁹⁷HS⁵⁹⁸CNE⁵⁹⁹AP⁶⁰⁰PL⁶⁰¹MA⁶⁰²FC⁶⁰³GH
RD⁶⁰⁴VNS⁶⁰⁵PEES⁶⁰⁶Q⁶⁰⁷RELL⁶⁰⁸F⁶⁰⁹HER⁶¹⁰LK⁶¹¹SAL⁶¹²DK⁶¹³TF⁶¹⁴RC⁶¹⁵SES⁶¹⁶QRAS⁶¹⁷YQ⁶¹⁸KL⁶¹⁹DAI⁶²⁰TEL⁶²¹Y⁶²²DP⁶²³Q⁶²⁴Q⁶²⁵INN⁶²⁶FM⁶²⁷DT⁶²⁸FK⁶²⁹ML⁶³⁰WD
Q⁶³¹ST⁶³²AVE⁶³³GD⁶³⁴AK⁶³⁵IL⁶³⁶RE⁶³⁷FF⁶³⁸Q⁶³⁹NLL⁶⁴⁰IH⁶⁴¹VP⁶⁴²IA⁶⁴³VT⁶⁴⁴KI⁶⁴⁵PVL⁶⁴⁶AN⁶⁴⁷V⁶⁴⁸LQ⁶⁴⁹LI⁶⁵⁰HA⁶⁵¹Q⁶⁵²ID⁶⁵³FL⁶⁵⁴Q⁶⁵⁵TS⁶⁵⁶ID⁶⁵⁷IF⁶⁵⁸Q⁶⁵⁹K⁶⁶⁰AT⁶⁶¹VE
LV⁶⁶²PR⁶⁶³RH⁶⁶⁴KT⁶⁶⁵WF⁶⁶⁶II⁶⁶⁷PI⁶⁶⁸IS⁶⁶⁹FL⁶⁷⁰K⁶⁷¹HM⁶⁷²IGIS⁶⁷³GV⁶⁷⁴AR⁶⁷⁵Q⁶⁷⁶K⁶⁷⁷HS⁶⁷⁸VS⁶⁷⁹Q⁶⁸⁰VL⁶⁸¹KE⁶⁸²VE⁶⁸³FR⁶⁸⁴CRET⁶⁸⁵FARD⁶⁸⁶YV⁶⁸⁷VEN⁶⁸⁸DN⁶⁸⁹VS⁶⁹⁰ID⁶⁹¹H
RG⁶⁹²AK⁶⁹³STAL⁶⁹⁴FAS⁶⁹⁵CY⁶⁹⁶NTNS⁶⁹⁷IR⁶⁹⁸Q⁶⁹⁹N⁷⁰⁰FL⁷⁰¹LL⁷⁰²VD⁷⁰³SE⁷⁰⁴AFI⁷⁰⁵K⁷⁰⁶EA⁷⁰⁷PT⁷⁰⁸IL⁷⁰⁹GF⁷¹⁰LAQ⁷¹¹NT⁷¹²K⁷¹³LI⁷¹⁴IS⁷¹⁵TNT⁷¹⁶TS⁷¹⁷DS⁷¹⁸TC⁷¹⁹FL
RL⁷²⁰NA⁷²¹NP⁷²²DM⁷²³AN⁷²⁴VVS⁷²⁵YCE⁷²⁶HH⁷²⁷SP⁷²⁸TE⁷²⁹KOD⁷³⁰ATAC⁷³¹CV⁷³²RL⁷³³HK⁷³⁴PT⁷³⁵SL⁷³⁶NS⁷³⁷Q⁷³⁸VRT⁷³⁹AN⁷⁴⁰MF⁷⁴¹MG⁷⁴²AM⁷⁴³DE⁷⁴⁴II⁷⁴⁵GOT
40 NK⁷⁴⁶IS⁷⁴⁷QNT⁷⁴⁸V⁷⁴⁹LT⁷⁵⁰Q⁷⁵¹SS⁷⁵²RE⁷⁵³FD⁷⁵⁴IL⁷⁵⁵RY⁷⁵⁶ST⁷⁵⁷INT⁷⁵⁸NAY⁷⁵⁹DP⁷⁶⁰GK⁷⁶¹LY⁷⁶²VY⁷⁶³LD⁷⁶⁴PAFT⁷⁶⁵TR⁷⁶⁶NR⁷⁶⁷KAS⁷⁶⁸GT⁷⁶⁹UAA⁷⁷⁰VAG⁷⁷¹Y⁷⁷²RH⁷⁷³Q⁷⁷⁴FLY⁷⁷⁵I
GLE⁷⁷⁶HF⁷⁷⁷RL⁷⁷⁸DS⁷⁷⁹SS⁷⁸⁰SA⁷⁸¹IA⁷⁸²BCAA⁷⁸³HM⁷⁸⁴IS⁷⁸⁵VL⁷⁸⁶SL⁷⁸⁷HP⁷⁸⁸Y⁷⁸⁹DEL⁷⁹⁰RI⁷⁹¹AVE⁷⁹²GN⁷⁹³Q⁷⁹⁴AA⁷⁹⁵VR⁷⁹⁶IA⁷⁹⁷CL⁷⁹⁸IR⁷⁹⁹Q⁸⁰⁰SV⁸⁰¹QS⁸⁰²ST⁸⁰³LI⁸⁰⁴R
VL⁸⁰⁵FY⁸⁰⁶HT⁸⁰⁷PD⁸⁰⁸NH⁸⁰⁹IE⁸¹⁰Q⁸¹¹FY⁸¹²LM⁸¹³GR⁸¹⁴KAL⁸¹⁵AVE⁸¹⁶Q⁸¹⁷FI⁸¹⁸SR⁸¹⁹FN⁸²⁰SY⁸²¹IKAS⁸²²QEL⁸²³VS⁸²⁴Y⁸²⁵TI⁸²⁶KL⁸²⁷SH⁸²⁸PI⁸²⁹YLL⁸³⁰EQ⁸³¹LI⁸³²Q⁸³³IL⁸³⁴RV
TL⁸³⁵AE⁸³⁶GT⁸³⁷ARY⁸³⁸SA⁸³⁹K⁸⁴⁰R⁸⁴¹NR⁸⁴²IS⁸⁴³DE⁸⁴⁴LI⁸⁴⁵IA⁸⁴⁶VIN⁸⁴⁷ATY⁸⁴⁸L⁸⁴⁹CD⁸⁵⁰DI⁸⁵¹HAIR⁸⁵²FR⁸⁵³VS
>gi_2746296
45 ML⁸⁵⁴RS⁸⁵⁵CD⁸⁵⁶IA⁸⁵⁷Q⁸⁵⁸KAY⁸⁵⁹QSI⁸⁶⁰W⁸⁶¹GE⁸⁶²Q⁸⁶³DK⁸⁶⁴LS⁸⁶⁵ST⁸⁶⁶FP⁸⁶⁷NSA⁸⁶⁸IF⁸⁶⁹Q⁸⁷⁰K⁸⁷¹FI⁸⁷²IT⁸⁷³PEL⁸⁷⁴GF⁸⁷⁵THAY⁸⁷⁶CR⁸⁷⁷HY⁸⁷⁸K⁸⁷⁹PL⁸⁸⁰Y⁸⁸¹LC⁸⁸²DR⁸⁸³Q⁸⁸⁴R
HVS⁸⁸⁵K⁸⁸⁶IA⁸⁸⁷IC⁸⁸⁸PL⁸⁸⁹CA⁸⁹⁰LS⁸⁹¹KL⁸⁹²FT⁸⁹³AI⁸⁹⁴IE⁸⁹⁵NT⁸⁹⁶EV⁸⁹⁷Q⁸⁹⁸Y⁸⁹⁹K⁹⁰⁰HL⁹⁰¹EL⁹⁰²Q⁹⁰³TS⁹⁰⁴FY⁹⁰⁵NR⁹⁰⁶MF⁹⁰⁷LQ⁹⁰⁸IE⁹⁰⁹K⁹¹⁰FI⁹¹¹Q⁹¹²DP⁹¹³RW⁹¹⁴IG⁹¹⁵GF⁹¹⁶NT

-22-

- NKKERIKLEPPQKRSILRIHIFFIISVTKLPTLANHVLDLYKYPFIDIEFINESVNILKQKASVFLVPRRHG
 KTWNFIVICFLKLNLEGISIGYVAHQKHVSFVMKDVEFKRRFFKRNITQDQNVITIEHETIKSTAL
 FASCYNTHSIRGQSFNLLIVDESHFIKDAFSTILGLPQSSSTRIIFISSTNSQNHSTSLTKLNSPFE
 MLTVSVYVCEDEHVHILNDRGNATTCACYRLHKPKFISINADVKKTADLFLKGAFKHEIMGGSLCNVNDIT
 5 LITEQGLIEFLDFRYSSTISKQIIPFLGKELYIYIDPAYTINRRASGTGVAAGTYIGDYITYGMEHYFLE
 SLGNSDASIAEACAHMILAVLELHPFFTEVKIIIEGNSNQASAVKIACILEQITISVIRYKHITTEFTFLD
 QSQIAQFFYLKGKELKLAVEFFISNPNNGYIKASQELISFTIKITYDPVEYIEQIKNLHQININEHVY
 NAKKQTCSDDLIIAIIIMAIYVCSGNSSASFREI
 >gi_125496
 10 MLKTCDTIHIKNYEAIIWKGERDCSTISTKYPSAIFYKKRFIMLTPELGFAHSYNQVKPLYTPCEKQ
 RHLKMRKFLTILPSLTKLQEMKFLPASDKSFESQYTEFLESFKILVREPLFLQIDGFIKDFRKNWKEGF
 NDFGDTFRKIQLPEFFQKNILIHVIFFLAVTKLALANRVINLTHVDFIEFVNESTLNTLQKTKNVFLVPR
 RHGKTWFIIVPIISFLNLEGISIGYVAHQKHVSFVMKEVEFKRRMFFKRTITCLDNVITIDHGNIKS
 TALFASCYNTHSIRGQSFNLLIVDESHFIKDAFSTILGLPQASTKILFISSTNSGNHSTSFLMKLNN
 15 SPFEMLSVSVYVCEDEHMLNERNGNATCSCYRLHKPKFISINAEVKKTANFLLEGAFIHEINGGATCNV
 INDVILTEQQTTEFFRYSSTINKNLIIFLQKDLVYLDPAYTGNRRASGTGIAAGTYIGDYITYGMEHY
 YFLESLMTSSDTAIAEACAHMILSILDLHPFFTEVKIIIEGNSNQASAVKIACIIKENITANKSIQVTF
 HTPDQNIQAQFFYLKGKELKLAVEFFISNPNNGYIKASQELISFTIKITYDPVEYIALEQIRNIHQISVNN
 YITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
 20 >gi_854039
 MKLNNSPFEMLSVSVYVCEDEHMLNERNGNATCSCYRLHKPKFISINAEVKKTANFLLEGAFIHEINGG
 ATCNVINDVILTEQQTTEFFRYSSTINKNLIIFLQKDLVYLDPAYTGNRRASGTGIAAGTYIGDYITYV
 YGMEHYFLESMTSSDTAIAEACAHMILSILDLHPFFTEVKIIIEGNSNQASAVKIACIIKENITANKSI
 QVTFHTPDQNIQAQFFYLKGKELKLAVEFFISNPNNGYIKASQELISFTIKITYDPVEYIALEQIRNIHQ
 25 ISVNNYITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
 >gi_5733564
 MLKTCDTIHIKNYEAIIWKGERNCSTISTKYPSAIFYKKRFIMLTPELGFAHSYNQVKPLYTPCEKQ
 RHLKMRKFLTILPSLTKLQEMKFLPASDKSFESQYTEFLESFKILVREPLFLQIDGFIKDFRKNWKEGF
 NDFGDTFRKIQLPEFFQKNILIHVIFFLAVTKLALANRVINLTHVDFIEFVNESTLNTLQKTKNVFLVPR
 30 RHGKTWFIIVPIISFLNLEGISIGYVAHQKHVSFVMKEVEFKRRMFFKRTITCLDNVITIDHGNIKS
 TALFASCYNTHSIRGQSFNLLIVDESHFIKDAFSTILGLPQASTKILFISSTNSGNHSTSFLMKLNS
 PFEMLSVSVYVCEDEHMLNERNGNATCSCYRLHKPKFISINAEVKKTANFLLEGAFIHEINGGATCNV
 INDVILTEQQTTEFFRYSSTINKNLIIFLQKDLVYLDPAYTGNRRASGTGIAAGTYIGDYITYGMEHY
 FLESMTSSDTAIAEACAHMILSILDLHPFFTEVKIIIEGNSNQASAVKIACIIKENITANKSIQVTFH
 35 TPDQNIQAQFFYLKGKELKLAVEFFISNPNNGYIKASQELISFTIKITYDPVEYIALEQIRNIHQISVNNY
 IITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
 >gi_4996048
 MKLNNSPFEMLSVSVYVCEDEHMLNERNGNATCSCYRLHKPKFISINAEVKKTANFLLEGAFIHEINGG
 ATCNVINDVILTEQQTTEFFRYSSTINKNLIIFLQKDLVYLDPAYTGNRRASGTGIAAGTYIGDYITYV
 40 YGMEHYFLESMTSSDTAIAEACAHMILSILDLHPFFTEVKIIIEGNSNQASAVKIACIIKENITANKSI
 QVTFHTPDQNIQAQFFYLKGKELKLAVEFFISNPNNGYIKASQELISFTIKITYDPVEYIALEQIRNIHQ
 ISVNNYITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
 >gi_1136608
 MLLSHNRFLANLEETAKDAGERNELSAPTFTTHCPTKPMARHPFIOVHRINSYSSVLETYCTRHHPA
 45 TPTSANPDVGTTPRSEDNVPAKTRLESLSLYLMRCVREDANHSADQLWEYQAGRKTHDSLHACSYYR
 ELQAFVNLNLSFLNGCYVPGVHWLEPPQQQLVNHITFFFLVSIKAPQKTEQLPGLFKYFGLFETPNSVLQ

-23-

- TFKQKASVFLIPRRHGKTIWVVAIISMLLASVENINIGYVAHQKHVANSVFAEIIKTLGRNFPFKMLNLIK
KENGITIIYTRPGGRSSSLNCACTFNHNSIRGQTFNLLYVDEANFIKDALPAILGFMLQKDAKLIPISSV
NSSDRSTSFLLNLRNAQEKMLANVVSVYVCAHREDFHLQDALVSCPCYRLHIPTYITIDESIKTTTNLFME
GAFDTLMGEGAASSNATLYRVVGDAALQTFDMCRVDTTAQQVQKCLGKQLFVYIDPAYTNTEASOTGV
5 GAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTIERVNAAVEGNSQDSGVA
IATVLNEICPLPIHFLHYTDKSSALQWPIYMLGEEKSAFETFIYALNSOTLSASQTVVSTIKISFDPV
TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDOLLVAVMAHFMATDDRHMYKPISPQ
>gi_1718281
MLQKDAKLIPISSSVNSSDRSTSFLLNLRNAQEKMLANVVSVYVCAHREDFHLQDALVSCPCYRLHIPTYIT
10 IDESIKTTTNLFMEGAFDTLMGEGAASSNATLYRVVGDAALQTFDMCRVDTTAQQVQKCLGKQLFVYID
PAYTNTEASOTGVGAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTIERVNA
AAVEGNSQDSGVAIATVLNEICPLPIHFLHYTDKSSALQWPIYMLGEEKSAFETFIYALNSOTLSASQ
TVVSTIKISFDPVTYLVEQVRAIKCVLEDDGGQSYSAKQKMSDOLLVAVMAHFMATDDRHMYKPISP
Q
15 >gi_2246515
MLQKDAKLIPISSSVNSSDRSTSFLLNLRNAQEKMLANVVSVYVCAHREDFHLQDALVSCPCYRLHIPTYIT
IDESIKTTTNLFMEGAFDTLMGEGAASSNATLYRVVGDAALQTFDMCRVDTTAQQVQKCLGKQLFVYID
PAYTNTEASOTGVGAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTIERVNA
AAVEGNSQDSGVAIATVLNEICPLPIHFLHYTDKSSALQWPIYMLGEEKSAFETFIYALNSOTLSASQ
20 TVVSTIKISFDPVTYLVEQVRAIKCVLEDDGGQSYSAKQKMSDOLLVAVMAHFMATDDRHMYKPISP
Q
>gi_2246552
MLLSHREERLAANLQETAJDAGERWLSAPFTIRHCPTARMARHFFIGVVRINSYSSVLETYCTRHPHA
TPTSANFDVOTPRPSEDNVPAKPELLESISTYLQNRQVREDARVSTADQLVEYQAARKTHDSLACSVYR
25 ELQAFVLNMLSSFLANGCYVFGVHWLSEFFQQQLVMHTEFFFLVS IKAPQKTHQLGLFLKQYGLFETPNSVLQ
TFKQKASVFLIPRRHGKTIWVVAIISMLLASVENINIGYVAHQKHVANSVFAEIIKTLGRNFPFKMLNLIK
KENGITIIYTRPGGRSSSLNCACTFNHNSIRGQTFNLLYVDEANFIKDALPAILGFMLQKDAKLIPISSV
NSSDRSTSFLLNLRNAQEKMLANVVSVYVCAHREDFHLQDALVSCPCYRLHIPTYITIDESIKTTTNLFME
30 GAFDTLMGEGAASSNATLYRVVGDAALQTFDMCRVDTTAQQVQKCLGKQLFVYIDPAYTNTEASOTGV
GAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTIERVNAAVEGNSQDSGVA
IATVLNEICPLPIHFLHYTDKSSALQWPIYMLGEEKSAFETFIYALNSOTLSASQTVVSTIKISFDPV
TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDOLLVAVMAHFMATDDRHMYKPISPQ
>gi_4494933
MLQKDAKLIPISSNSSDKSTSFLLNLKDAHEKMLANVVSVYVCPDHKDDFNLDQTVVACPCYRLHI PAYIT
35 IDSTVRSSTNLFLEGAFTSLMGDAANTSQAQSMHKIVSDSSLSQDLCRVESTSQDIQGANKEPCLHVIYDP
AYTNTEASOTGIGAVIAVNHKVIKILLGVEHFFLRDLTGTAAYQLASCAALIRAIATVLHPTIHRVNA
AVEGNSQDQAGVIAIATVLNEICSVPLSFLHHVDKNTLIRSFIYMLGPEKAKAFESFIYALNSGTFSSASQT
VVSTIKISFDPVAYLIDQIKAIRICPLKDGGHYCAKQKTMSSDVLVAAVMAHFMATNDKFFVESKLS
>gi_7330018
40 MLQKDAKLIPISSNSSDKSTSFLLNLKDAHEKMLANVVSVYVCPDHKDDFNLDQTVVACPCYRLHI PAYIT
IDSTVRSSTNLFLEGAFTSLMGDAANTSQAQSMHKIVSDSSLSQDLCRVESTSQDIQGANKEPCLHVIYDP
AYTNTEASOTGIGAVIAVNHKVIKILLGVEHFFLRDLTGTAAYQLASCAALIRAIATVLHPTIHRVNA
AVEGNSQDQAGVIAIATVLNEICSVPLSFLHHVDKNTLIRSFIYMLGPEKAKAFESFIYALNSGTFSSASQT
45 VVSTIKISFDPVAYLIDQIKAIRICPLKDGGHYCAKQKTMSSDVLVAAVMAHFMATNDKFFVESKLS
>gi_4019255
MLLLEAKALMENLTEASTQSETWTVDTPTIMINIKESRMAYSIGKIVIPSNLYASLITSPCRLYRP

-24-

LALKQPLFQTGTLRALPSEKPYISQKLSNYKSLTLKHVHMDTEAEAYASVQTEKTFMECPYILELRQ
 FNLNSSFNGCYVKRSTHIEPFQQLILHTFYFLISIKSPSTNKLDFIPKEYFQLGEMDSAMLQNFQKQ
 KASIFLIPRRHGKTWIVVAIISMLTISVENLHVGVVARQKHVANSVFTEIINTLQKWFPSENIDVKKENG
 TIIYKIPGKPSITLMCASCFTNRSIRGQTFNLLYIDEANFIKSDLPAILGFMQLQKDAKLIPISSVNSGD
 5 KATSFLNLRKASEKMLNINVIICPDHKKDFFSLQDSLISCPCYKLYIPTVITIDETIKNTNLFLOGAFT
 TELMGDISVMSKNNIHKVIGETALMQFDLCRIDTTKPEITQCLNSIMVLYIDPAYTNSEASGTGIGAI
 ALNENSSKCIIVGIEHYFLKDLTGATYQIASCACSLIRAAVLVYPHIQAVHVAVEGNSQDSAVAISTF
 INECSPVKNVPMHYKDKTTAMQWPIYNLGSSEKSAQFESFIYAINSOTISASQSIIISNTIKLTFDPIISYLI
 EQRAIRCYPLRDGSHTYCAKERTVSDDVLVAVVMHFFSTSNKHIFKQLNSI
 10 >gi_4019257
 MLQKDAKLIPISSVNSGDKATSFLNLRKASEKMLNINVIICPDHKKDFFSLQDSLISCPCYKLYIPTVIT
 IDETIKNTNLFLOGAFTTELMGDISVMSKNNIHKVIGETALMQFDLCRIDTTKPEITQCLNSIMVLYID
 PAYTNSEASGTGIGAIILAKNSKCIIVGIEHYFLKDLTGATYQIASCACSLIRAAVLVYPHIQAVH
 VAVEGNSQDSAVAISTFNECSPVKNVPMHYKDKTTAMQWPIYNLGSSEKSAQFESFIYAINSOTISASQ
 15 SIIISNTIKLTFDPIISYLIQIRAIRCYPLRDGSHTYCAKERTVSDDVLVAVVMHFFSTSNKHIFKQLNSI
 I
 >gi_60355
 MLLLKAKKAIENLSEVSSTQAEITDNDMSTPIIITNTSISERTAYSIGVIPSVNLYSSTLTSFCKLHP
 LTLNQTQPGTGTALRLPHEKPLILQDLSNYKLLTSQNVCHDTEANTEYNAVQQTORTSCMECPYILELRQ
 20 FVNLNSSFNGCYVKRSTHIEPFQQLILHTFYFLISIKSPSTNKLDFIPKEYFQLGEMDPMQLQIFKQ
 KASIFLIPRRHGKTWIVVAIISMLTISVENLHVGVVARQKHVANSVFTEIINTLQKWFPSEYIDIKKENG
 TIIYKIPGKPSITLMCASCFTNRSIRGQTFNLLYIDEANFIKSDLPAILGFMQLQKDAKLIPISSVNSGD
 RATSFLNLRKASEKMLNINVIICPDHKKDFFSLQDSLISCPCYKLYIPTVITIDETIKNTNLFLOGAFT
 TELMGDSGISKSNHKKVISEMAITQFDLCRADTTKPEITQCLNSIMVLYIDPAYTNSEASGTGIGAIL
 25 TFKNSKCIIVGIEHYFLKDLTGATYQIASCACSLIRAAVLVYPHIQAVHVAVEGNSQDSAVAISTF
 INECSPIKVVFIHYKDKTTAMQWPIYNLGAESKIAFESFIYAINSOTISASQSIIISNTIKLTFDPIISYLI
 EQRSIRCYPLRDGSHTYCAKERTVSDDVLVAVVMHFFSTSNKHIFKPLNST
 >gi_595201
 MLQKDAKLIPISSVNSDQTSFLYNLRKAKEMLNINVIICPDHREDFSLOESVVS CPCYELHIPTVIA
 IDENIKDTNLFMEGAFTTELMGDGAATTPQNNHKKVVOEPALVQFDLCRVDTGSPEAQGLNFTFLYV
 30 DPAYTNTEASGTGIGAVVSMKNSDRCVVGVVEHFPLKELTGASSLIQASCAALIRSLATLHPFVREAH
 VAEGNSQDSAVAISTLHRSPLPVKFLHDAKATGVQNMVYILGAEKARAFETFIYALNSPTLSQCG
 AIVSNTIKLSFDPVAVLIEQIRAIKCYFLKDGTVSYCAKHGGSDTLVAVVMHFFSTSNKHIFKPLNST
 QI
 35 >gi_4928934
 MLLSSFRNLQNKYKYSVOAQNIDWFVETPVLISKDSKTNRLAHPILGVISRINLYSPTLKYCYEYST
 TKQPKTFPDIGVVRDLKKHDQYFLPKLQHLSTLCEAHVHVDRAQAVEFNAILTKAFHANGVLNMLKQ
 FLNLNSSFNGCYVSKSTCIELFQKQLILHTFYFLISIKPSTNKLMTFFFKHYVGLFDIDNNLQCFKQ
 KSTVFLIPRRHGKTWIVVAIISVLLASVENVHIGYVAHQKHVANSVFTEIINTLQKWFPSEKIEIKKENG
 40 TIIYKIPGKPSITLMCASCFTNRSIRGQTFNLLYIDEANFIKKEALPAILGFMQLQKDAKLIPISSVNSAD
 KSTSLFNLRKASEKMLNINVIICPDHKKDFFSLQDSLISCPCYKLYIPTVITIDESIKNTNLFLOGDVT
 TELMGDISFTFSMFKVVEEQALFHFDICRVDTQIDTVKIIDNVLVYVDPAYTNSEASGTGIGAVV
 PLKTKVXTIILGIEHYFLKDLTGATYQIASQYAYCVTSMIKAILITHPHINHVNVVAEGNSQDSAVAISTFI
 KEYCVPVVFPAKNGSSVFWQPIYILGSEKSAQPEKFCALITQTLGASQTVISNTIKLSFDPVAVLIEQ
 45 QIRAIRCLPLRDGSHTYCAKQNTSDTLVAVVMHFFSTSNKHIFKPELCKT
 >gi_1632798

-25-

MLYASQGRGLTENLQDQDSTTQCLGASTPSIMYTGAKSDRWAPHLVGTIHAENLYCPMLRAYCRHY
 GFRPVFVASDESILWFGASPALHTFPVQVQMLPELEDTLQRLPPNLEDSALTEFKTSVSSARILE
 DPNFLEMREFVTSLASFLSQYKHPARLEAFQKQVLSHFYFLISIKSLEITDITMFDIPQSAFGLSEMT
 LEKLHIFKQKASVFLIPRRHOKTWIVVAIISLILSNLSNVQIGYVAHQKHVAAVFTEIIDTLTKSFDSK
 5 RVEVNKSTTITFRHSQKISSTVMCATCFNKNISIRGQTVHLLFVDEANFIKKEALPAILGPMQKDAKII
 FISSVNSADQATSFILYKLDQERILNVVSYVCQERQDFDQDQSMVSCPCFRLHIFSYITMDSENTRATT
 NLFLDGAFSTELMGDTSSLSQGSLSRTVRDDANQJELCRVDTLNFRVAGRLASSLYVYVDFAVTRNPSA
 SGTGIAAVTHDRADPNRVIVLGHEHFFLKDLTGDAALQIATCVVALVSSIIVTLRPHLEEVKVAVEGNSSQ
 DSAVAIASIIGESCELEPCAFVHTKDKTSSLQWPMVLLTNEKSKAFERLIYAVMTASLASAQVTVSNTIQL
 10 SDFDVLVLISQIRAKIPILRDGTYTYTGKQENLSDVVALVMAHFLAITTQKHYPKKVH
 >gi_2337931
 MFTVKWMPALQKACEELQWQSAKSGKNFPVPEFTPLVAVETRSESNWHPYLGLLPQVAAYSSTLEDYCHL
 YNPYIDALTRCDLQTHRVAQPVLSQDLQQLKKLSPCEPNTSVKAKLEFAAVTRHQALDNSQVLE
 LKTFVLNLSAFLNKRYSDBRSHIELFQKQLINHTFFFLVSIIKAPELCEKFCNIFKLYFNIDTMDQATLDI
 15 FEQKASVFLIPRRHOKTWIVVAIISILLASVQDLRIQYVAHQKHVAAVFTEVINTLHFFPGKMDVKK
 ENGTIIIFGLPNKFPSTLLCATCFNKNISIRGQTVHLLFVDEANFIKKEALPTLIGPMQKDAKIIFISSSN
 SSDQSTSFILYNLKGASERMLNVVSYVCSNKKEDFSMQGLISCPYSLHVSYSIIDBQIKTTNLLFDG
 VFDTELMGDSSCOTLSTFQIISSEALSQFELCRIDTASPVQAHLNSTVMYIDPAFTNNLDASGTGISV
 IGLGAKTKVILQCEHFFLQKLTGTAALQIASCATSLRSVVIIEHMIKCAQITIBGNSQDSAVAIANF
 20 IDECAPIPVTFYHQSDKTKGVLCFLYLLGQEKAVAFESFIYAMNGLCKASQLIVSHRTIKLSDFPVTYLL
 BQVRAIKQCSLRDGSHTYHAKQENLSDDLLVSVMSLYLSANTLFFKPLHIERFF
 >gi_2317977
 MLQKDAKIFFISSVMSGKTTSFILYNLKDANEKMNWVVSVCSEHMEDEFNKSAITACPCYRLVPEFIT
 INDNKCTNNLLEGSFATELMGNMQSHTEVSQNSMIHESSLTRDFYRCDTAQGAPPTENTLFWYIDF
 25 AYGNMVHVASGTGIVAMSHCKHTKKCIILGLEHFFLANLTGTAARNIASCATALLEGILFQHPWQIQRICI
 IEGNSNQDSAVAIAITFISHNIKPLTFASFYRDKTMQWPIYMLSGDKTLAFQNFISSLNQGLLCAQTVV
 SNTVLLSSDPISYLBQIKNTKCIYHKNKTIITFQSKTHMTMSDDVLIACVMTCYVMTNKISYISFSIK
 >gi_6625593
 MFIAKSKSVFAVYRSTVSSHSEEPNKSDDPVYPTQYKKQCNRLPNAYLOTLEHASKYSENFRHYVATFS
 30 NSFLDFFQSVFNERNFCESYVPLYDSALQCSAKTLVGCSVSTTERNEYEVCKEATRCFKDAMSHKVLKVF
 LSLNFWLKGHYKSAQFALEFFQKQLILHSMFNVASIKCPETTTKLDFEFLLDMLYFDNLDLLFLQK
 SPAPLIPRRHOKTWIVTAIISMLTSDVDLHIGYVAHQKHVSLAVFLEISWILLANFPRNRIDIKKENG
 ILYSLPBGKSSSTMATCFNKNISIRGQTVHLLFVDEANFIKKEALPAILGPMQKDAKIFFISSVMSGK
 TTSFILYNLKDANEKMNWVVSVCSEHMEDEFNKSAITACPCYRLVPEFITINDNIKCTNNLLEGSFAT
 35 ELANGMQSHTEVSQNSMIHESSLTRDFYRCDTAQGAPPTENTLFWYIDFAYGNMVHVASGTGIVAMSH
 CKHTKKCIILGLEHFFLANLTGTAARNIASCATALLEGILFQHPWQIQRICIIEGNSNQDSAVAIAITFISH
 NIKLPLTFASFYRDKTMQWPIYMLSGDKTLAFQNFISSLNQGLLCAQTVVNTVLLSSDPISYLBQIK
 NTKCIYHKNKTIITFQSKTHMTMSDDVLIACVMTCYVMTNKISYISFSIK

40

131	140	150	160	170	180	190	200	210	220	230	240	250	260
1	2	3	4	5	6	7	8	9	10	11	12	13	14
15	16	17	18	19	20	21	22	23	24	25	26	27	28
29	30	31	32	33	34	35	36	37	38	39	40	41	42
43	44	45	46	47	48	49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80	81	82	83	84
85	86	87	88	89	90	91	92	93	94	95	96	97	98
99	100	101	102	103	104	105	106	107	108	109	110	111	112
113	114	115	116	117	118	119	120	121	122	123	124	125	126
127	128	129	130	131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150	151	152	153	154
155	156	157	158	159	160	161	162	163	164	165	166	167	168
169	170	171	172	173	174	175	176	177	178	179	180	181	182
183	184	185	186	187	188	189	190	191	192	193	194	195	196
197	198	199	200	201	202	203	204	205	206	207	208	209	210
211	212	213	214	215	216	217	218	219	220	221	222	223	224
225	226	227	228	229	230	231	232	233	234	235	236	237	238
239	240	241	242	243	244	245	246	247	248	249	250	251	252
253	254	255	256	257	258	259	260	261	262	263	264	265	266
267	268	269	270	271	272	273	274	275	276	277	278	279	280
281	282	283	284	285	286	287	288	289	290	291	292	293	294
295	296	297	298	299	300	301	302	303	304	305	306	307	308
309	310	311	312	313	314	315	316	317	318	319	320	321	322
323	324	325	326	327	328	329	330	331	332	333	334	335	336
337	338	339	340	341	342	343	344	345	346	347	348	349	350
351	352	353	354	355	356	357	358	359	360	361	362	363	364
365	366	367	368	369	370	371	372	373	374	375	376	377	378
379	380	381	382	383	384	385	386	387	388	389	390	391	392
393	394	395	396	397	398	399	400	401	402	403	404	405	406
407	408	409	410	411	412	413	414	415	416	417	418	419	420
421	422	423	424	425	426	427	428	429	430	431	432	433	434
435	436	437	438	439	440	441	442	443	444	445	446	447	448
449	450	451	452	453	454	455	456	457	458	459	460	461	462
463	464	465	466	467	468	469	470	471	472	473	474	475	476
477	478	479	480	481	482	483	484	485	486	487	488	489	490
491	492	493	494	495	496	497	498	499	500	501	502	503	504
505	506	507	508	509	510	511	512	513	514	515	516	517	518
519	520	521	522	523	524	525	526	527	528	529	530	531	532
533	534	535	536	537	538	539	540	541	542	543	544	545	546
547	548	549	550	551	552	553	554	555	556	557	558	559	560
561	562	563	564	565	566	567	568	569	570	571	572	573	574
575	576	577	578	579	580	581	582	583	584	585	586	587	588
589	590	591	592	593	594	595	596	597	598	599	600	601	602
603	604	605	606	607	608	609	610	611	612	613	614	615	616
617	618	619	620	621	622	623	624	625	626	627	628	629	630
631	632	633	634	635	636	637	638	639	640	641	642	643	644
645	646	647	648	649	650	651	652	653	654	655	656	657	658
659	660	661	662	663	664	665	666	667	668	669	670	671	672
673	674	675	676	677	678	679	680	681	682	683	684	685	686
687	688	689	690	691	692	693	694	695	696	697	698	699	700
701	702	703	704	705	706	707	708	709	710	711	712	713	714
715	716	717	718	719	720	721	722	723	724	725	726	727	728
729	730	731	732	733	734	735	736	737	738	739	740	741	742
743	744	745	746	747	748	749	750	751	752	753	754	755	756
757	758	759	760	761	762	763	764	765	766	767	768	769	770
771	772	773	774	775	776	777	778	779	780	781	782	783	784
785	786	787	788	789	790	791	792	793	794	795	796	797	798
799	800	801	802	803	804	805	806	807	808	809	810	811	812
813	814	815	816	817	818	819	820	821	822	823	824	825	826
827	828	829	830	831	832	833	834	835	836	837	838	839	840
841	842	843	844	845	846	847	848	849	850	851	852	853	854
855	856	857	858	859	860	861	862	863	864	865	866	867	868
869	870	871	872	873	874	875	876	877	878	879	880	881	882
883	884	885	886	887	888	889	890	891	892	893	894	895	896
897	898	899	900	901	902	903	904	905	906	907	908	909	910
911	912	913	914	915	916	917	918	919	920	921	922	923	924
925	926	927	928	929	930	931	932	933	934	935	936	937	938
939	940	941	942	943	944	945	946	947	948	949	950	951	952
953	954	955	956	957	958	959	960	961	962	963	964	965	966
967	968	969	970	971	972	973	974	975	976	977	978	979	980
981	982	983	984	985	986	987	988	989	990	991	992	993	994
995	996	997	998	999	1000	1001	1002	1003	1004	1005	1006	1007	1008
1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020	1021	1022
1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036
1037	1038	1039	1040	1041	1042	1043	1044	1045	1046	1047	1048	1049	1050
1051	1052	1053	1054	1055	1056	1057	1058	1059	1060	1061	1062	1063	1064
1065	1066	1067	1068	1069	1070	1071	1072	1073	1074	1075	1076	1077	1078
1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092
1093	1094	1095	1096	1097	1098	1099	1100	1101	1102	1103	1104	1105	1106
1107	1108	1109	1110	1111	1112	1113	1114	1115	1116	1117	1118	1119	1120
1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134
1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148
1149	1150	1151	1152	1153	1154	1155	1156	1157	1158	1159	1160	1161	1162
1163	1164	1165	1166	1167	1168	1169	1170	1171	1172	1173	1174	1175	1176
1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190
1191	1192	1193	1194	1195	1196	1197	1198	1199	1200	1201	1202	1203	1204
1205	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218
1219	1220	1221	1222	1223	1224	1225	1226	1227	1228	1229	1230	1231	1232
1233	1234	1235	1236	1237	1238	1239	1240	1241	1242	1243	1244	1245	1246
1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260
1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274
1275	1276	1277	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288
1289	1290	1291	1292	1293	1294	1295	1296	1297	1298	1299	1300	1301	1302
1303	1304	1305	1306	1307	1308	1309	1310	1311	1312	1313	1314	1315	1316
1317	1318	1319	1320	1321	1322	1323	1324	1325	1326	1327	1328	1329	1330
1331	1332	1333	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344
1345	1346	1347	1348	1349	1350	1351	1352	1353	1354	1355	1356	1357	1358
1359	1360	1361	1362	1363	1364	1365	1366	1367	1368	1369	1370	1371	1372
1373	1374	1375	1376	1377	1378	1379	1380	1381	1382	1383	1384	1385	1386
1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400
1401	1402	1403											

01_10180715	1	261	270	280	290	300	310	320	330	340	350	360	370	380	390
01_26731189	2	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ	ERLWVFEISQ
01_56839285	3	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI	ERLLRVFDI
01_18638337	4	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL	ATFLR VFEJPL
01_59501	5	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL	ATFLRVFEJPL
01_320232	6	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES	GHLKXWNTFES
01_371317	7	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES	GHLRWNTFES
01_5863808	8	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS	UCLRTVNFIPS
01_5708110	9	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL	NRYSVTNFISL
01_31370	10	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE
01_325486	11	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE	INYLTVNFIDE
01_5733554	12	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL
01_2746552	13	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL
01_1632798	14	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL
01_7330618	15	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL
01_4019257	16	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL
01_15507	17	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL
01_2317077	18	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL
01_054039	19	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL
01_4956048	20	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL
Consensus	21	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL	FBLFKQYFBL

TABLE 2 CONTINUED

391	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	520
g1_10180719	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	510
g1_7571169	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	500
g1_11983837	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	490
g1_595014	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	480
g1_330792	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	470
g1_373137	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	460
g1_5704110	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	450
g1_248296	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	440
g1_1813970	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	430
g1_2748296	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	420
g1_325496	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	410
g1_1136508	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	400
g1_4019295	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	391
g1_50365	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	381
g1_2337934	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	371
g1_1332790	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	361
g1_6225593	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	351
g1_7116881	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	341
g1_2268315	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	331
g1_7330633	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	321
g1_4013227	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	311
g1_8952921	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	301
g1_2317377	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	291
g1_4956048	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	281
Consensus	ENFIPRQVQVITIGL	LNQTKLIFVSSINSGKASTSELYELGSRDOLLNWTYVTCDEHKNKIVTDTNATSCSYLNKWPFTTHGGRHARTDEHFLPDSFNEITIGG	271

TABLE 2 CONTINUED

[illegible]

TABLE 2 CONTINUED

	781	791
	-----+	
gi_10180719		
gi_7673189		
gi_5689285		VERTICRPTER
gi_1863837	TS	
gi_53501	VS	
gi_2605592	Q	
gi_330752	Q	
gi_571317	RPPHD	
gi_5865808		
gi_5709110		
gi_6163566		
gi_2746296		
gi_325486		
gi_5733564		
gi_1136808		
gi_2246552		
gi_4019255		
gi_80355		
gi_4928934		
gi_2337591	FF	
gi_1632798		
gi_6655593		
gi_1718281		
gi_248515		
gi_498933		
gi_7330010		
gi_4018957		
gi_695201		
gi_2317577		
gi_854039		
gi_4996048		
Consensus		

TABLE 2 CONTINUED

Table 3. Degenerate primers generated by CODEHOP

Block x7263xbliD

T L Y V Y I D P
 oligo:5'-RACCTGTACGTGtayatngaycc-3' degen=64 temp=33.4 Extend clamp

T L Y V Y I D P A
 oligo:5'-RACCTGTACGTGTACatngayccngc-3' degen=128 temp=36.0 Extend clamp

T L Y V Y I D P A Y
 oligo:5'-RACCTGTACGTGTACATngayccngcnt-3' degen=128 temp=42.5 Extend clamp

Complement of Block x7263xbliD

Y I D P A Y T N N T
 atrnancrtggGCGGATGTGGTGTGTGT
 degen=64 temp=62.9 oligo:5'-TGTTGTGGTGTAGGCGggrtcnaarta-3'

D P A Y T N N T R A
 anctrggcngnaTGTGGTGTGTGGTGTGGTCCG
 degen=128 temp=61.8 oligo:5'-GCCTGGGTGTGTGTGGTGTangcnggrtcna-3'

D P A Y T N N T R A
 ctrgcnccnawGTGGTGTGTGTGGTGTGGTCCG
 degen=64 temp=61.0 oligo:5'-GCCTGGGTGTGTGTGGTGTgwanccnggrtc-3'

Block x7263xbliE

C I I F G M E H F F
 oligo:5'-TGGATCATCTTGGGATngarcaytwyt-3' degen=64 temp=55.7 Extend clamp

I F G M E H F F F L
 oligo:5'-CATCTTCGGCATGGAGcaytwytwyt-3' degen=64 temp=62.0

Complement of Block x7263xbliE

E H F F L R D L T G
 ctygtawrawGGACTTCCTGGACTGCCCG
 degen=32 temp=61.7 oligo:5'-CCCGTCAGGTCTCTCAGGwarwartgytc-3'

H F F L R D L T G
 tygtawrawrxaCTTCCTGGACTGCCCG
 degen=128 temp=60.8 oligo:5'-GCCCGTCAGGTCTCTCarrwarwartgyt-3'

H F F L R D L T G
 gtrawrawrxaCTTCCTGGACTGCCCG
 degen=60.8 oligo:5'-GCCCGTCAGGTCTCTCarrwarwartgyt-3' degen=64

Block x7263xbliF

E V H I A V E G N
 oligo:5'-GGACGTGCACGTGGCCrtngarggnaa-3' degen=64 temp=63.8

Complement of Block x7263xbliF

E G N S S Q D S A
 anctycnttrwGGTGTGGTCTGAGCGCGG
 degen=128 temp=62.7 oligo:5'-GGCGGAGTCCTGGTGGwrttnccytcna-3'

E G N S S Q D S A V
 ctyccnttrwsGTGGTGTGGTCTGAGCGCGG
 degen=64 temp=63.9 oligo:5'-CGGCGGAGTCCTGGTGTGswrttnccytc-3'

CLAIMS

1. A high throughput method for screening a biological sample for unknown viruses,
which method comprises
- 5 (a) subjecting DNA from the sample to PCR amplification conditions using
simultaneously multiple pairs of degenerate primers, wherein each primer binds a
sequence that is conserved across members of a family of viruses and each pair of
primers selectively directs amplification of sequence of said family;
- 10 (b) sequencing PCR product obtained in step (a); and
- (c) comparing the sequence of the PCR product with the sequences in at least one
database comprising viral sequences to determine whether the sequence is present in,
15 or absent from, the database, wherein absence of the sequence from the database
suggests that the sequence may be from an unknown virus.
2. A method according to claim 1 wherein from 12 to 300 pairs of degenerate
primers are used simultaneously.
- 20 3. A method according to claim 1 wherein from 24 to 96 pairs of degenerate primers
are used simultaneously.
4. A method according to claim 1, 2 or 3 wherein the PCR reaction step is carried
25 out in a multi-well plate.
5. A method according to claim 4 wherein the multi-well plate is a 96-well or 384-
well plate.
- 30 6. A method according to claim 4 or 5 wherein each well contains more than one
pair of the degenerate primers.

7. A method according to claim 6 wherein each well contains three pairs of the degenerate primers.
8. A method according to claim 6 or 7 wherein each pair of the primers used in the same well as other pair(s) of primers generates a PCR product of a different size to said other pair(s).
9. A method according to any one of claims 6 to 8 wherein each pair of primers used in the same well as other pair(s) of primers carries a different label from said other pair(s).
10. A method according to claim 9 wherein each pair of primers used in the same well as other pair(s) of primers carries a differently-coloured fluorescent label from said other pair(s).
11. A method according to any one of the preceding claims wherein multiple biological samples are screened by simultaneously subjecting DNA from the multiple samples to PCR reaction conditions in step (a) using simultaneously the multiple pairs of degenerate primers on the DNA from each sample.
12. A method according to claim 11 wherein from 2 to 80 samples are screened simultaneously.
13. A method according to claim 12 wherein from 5 to 40 samples are screened simultaneously.
14. A method according to any one of the preceding claims wherein DNA of multiple biological samples is mixed together to produce one or more pooled samples and the PCR reaction of step (a) is carried out on the pooled samples.
15. A method according to any one of the preceding claims wherein the DNA is

genomic DNA.

16. A method according to any one of claims 1 to 14 wherein the DNA is cDNA.

17. A method according to any one of the preceding claims wherein the multiple pairs of degenerate primers are designed by

(i) providing a plurality of amino acid sequences from members of a first virus family,

10

(ii) comparing the sequences to identify conserved regions,

(iii) designing a first primer pair using a computer based method, wherein each primer in the pair binds a nucleotide sequence that encodes a conserved region

15 identified in (ii) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the nucleotide sequences that encode conserved regions in members of the first virus family, and

(iv) repeating steps (i) to (iii) for each virus family.

20

18. A method according to any one the preceding claims wherein the biological sample is a human sample.

19. A method according to any one of the preceding claims which further comprises
25 determining whether the sequence of the PCR product is a sequence of human DNA.

20. A method according to any one of the preceding claims wherein, if the sequence of the PCR product is absent from the database comprising viral sequences, DNA walking is carried out to determine any sequence which flanks the sequence of the
30 PCR product.

21. A method according to any one of the preceding claims wherein, if the sequence of the PCR product is absent from the database comprising viral sequences, any virus comprising the sequence is isolated.
- 5 22. A method according to any one of the preceding claims which further comprises determining whether the sequence of the PCR product or sequence contiguous thereto is present in a specimen from a patient who has a disease, wherein any presence of a said sequence in the specimen suggests that the sequence may be from a virus which is causing or contributing to the disease.
- 10 23. A method according to any one of the preceding claims which further comprises obtaining a specimen from each member of a group of subjects with a disease;
- 15 determining whether the sequence of the PCR product or sequence contiguous thereto is present in the specimen from each member of the group; and determining whether the proportion of subjects in whom a said sequence is present is greater in the group of subjects who have the disease than in a control group of subjects who do not have the disease, wherein a said greater proportion suggests that the sequence may be from a virus which causes or contributes to the
- 20 disease.

FIG 1

1/1

